

CRFE

renewed 5/19/03  
SEARCH REQUEST FORM

Scientific and Technical Information Center

Access DB#

93097

5/1 138

5/8 112

Requester's Full Name:

David Fox

Examiner #: 65401

Date:

5/14/03

Art Unit: 1638

Phone Number 308-0280

Serial Number: 09/940,550

Mail Box and Bldg/Room Location: CM 19E15 Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc., if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Point of Contact:

Toby Port

Technical Info. Specialist

CM1 6A04

703-308-2527

Title of Invention:

PLEASE SCAN ONLY

Inventors (please provide full names):

Earliest Priority Filing Date:

8/00

\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

please do SEQUENCE SEARCH

for SEQ ID NOS 9 + 10

(native + mutant form of same gene)

NO TIMELY HITS

THANK YOU

- amendment due 5/31 -

STAFF USE ONLY

Type of Search

Vendors and cost where applicable

Searcher:

NA Sequence (#) 2

STN

Searcher Phone #:

AA Sequence (#)

Dialog

Searcher Location:

Structure (#)

Questel/Orbit

Date Searcher Picked Up: 5/6

Bibliographic

Dr. Link

Date Completed: 5/12

Litigation

Lexis/Nexis

Searcher Prep & Review Time: 12

Fulltext

Sequence Systems

Clerical Prep Time:

Patent Family

WWW/Internet

Online Time: 12

Other

Other (specify)

***This Page Blank (uspto)***

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 15:44:18 ; Search time 5977.5 seconds

(without alignments)  
9688.381 Million cell updates/sec

Title: US-09-940-550A-9

Perfect score: 2031  
Sequence: 1 atggcacaagggtgtgtgac.....cgggaagacgttagcgcgtag 2031Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

GenEmbl:\*

1: gb\_da:\*

2: gb\_hlg:\*

3: gb\_in:\*

4: gb\_ov:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sts:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vl:\*

15: em\_da:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_or:\*

21: em\_ov:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_sts:\*

28: em\_un:\*

29: em\_vl:\*

30: em\_hlg\_hum:\*

31: em\_hlg\_inv:\*

32: em\_hlg\_other:\*

33: em\_hlg\_mus:\*

34: em\_hlg\_pln:\*

35: em\_hlg\_rtd:\*

36: em\_hlg\_mam:\*

37: em\_hlg\_vrt:\*

38: em\_hlg\_hum:\*

39: em\_hlg\_mus:\*

40: em\_hlg\_mus:\*

41: em\_hlg\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2031	100.0	2031	6	AX394182
2	2015	99.2	2031	6	AX394183
3	2014	99.2	9901	6	AX394184
4	1812.6	89.2	2200	7	C31INDNA
5	1812.6	89.2	3711	8	AX114861
6	1812.6	89.2	5715	12	PCL414670
7	1812.6	89.2	41489	7	BPH6589
8	200.4	9.9	5967	12	EVU84006
9	198.6	9.8	912	6	AX259239
10	198.6	9.8	1577	6	AX259244
11	198.6	9.8	1618	6	AX259242
12	198.6	9.8	2943	6	AX259250
13	195	9.6	2001	6	AX259240
14	195	9.6	2001	6	AX356863
15	195	9.6	2633	6	AR037156
16	195	9.6	2633	6	AR112042
17	195	9.6	2725	6	AX259245
18	195	9.6	2730	6	AX259243
19	195	9.6	4072	6	AX259251
20	195	9.6	4341	6	AX259249
21	195	9.6	5288	12	ASTDNABV
22	195	9.6	5534	6	AR037157
23	195	9.6	5534	6	AR112043
24	195	9.6	12817	6	AX356864
25	195	9.6	13274	6	AX356862
26	194	9.6	1701	6	AR174855
27	194	9.6	3336	6	AR174856
28	194	9.6	3877	6	AR174858
29	190.2	9.4	5560	6	AE01112
30	190.2	9.4	1034	6	AR098311
31	190	9.4	1034	6	AR87679
32	181.6	8.9	363	6	AX060392
33	180.6	8.9	363	6	AX060389
34	180.6	8.9	363	6	AX060391
35	180.6	8.9	369	6	AX060390
36	179.8	8.9	3699	8	STRSIG
37	64	3.2	125020	9	AR429315
38	63.8	3.1	139514	1	SGR300302
39	60.4	3.0	125020	9	AF429315
40	59	2.9	134375	2	AC118169
41	58.6	2.9	169546	2	AC004157
42	58.2	2.9	154902	9	AC005799
43	58.2	2.9	191947	2	AC079210
44	58	2.9	9424	8	SD0404228
45	57.8	2.8	1181	8	MISCOX13

## ALIGNMENTS

RESULT 1	AX394182	2031 bp	DNA	Linear	PAT 23-MAR-2002
LOCUS	AX394182				
DEFINITION	Sequence 9 from Patent WO0216609.				
ACCESSION	AX394182				
VERSION	AX394182.1	GI:19702116			
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS	Manikin, L.				
TITLE	Self-excising polynucleotides and uses thereof				
JOURNAL	Patent: WO 0216609-A 9 28-FEB-2002				
	Basf Plant Science GmbH (DE) ; McKersie, Bryan (US)				

filed at  
instat US

```

/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Clint DNA sequence"
451 a      562 c      640 g      378 t
BASE COUNT
ORIGIN

```

Query March	100.0%	Score 2031;	DB 6;	Length 2031;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 2031; Conservative	0;	Mismatch	0;	

[illegible]

Db	901	CCCTTAAAGCCCGGGACGTCAAGCCGCCATTCACC	CGGGACGATCAACGGGGCTTTGTAG	96	
QY	961	CGCATGAGCGTGTAGACCGGTGCGCAACCGGGGGGAGACGAT	TTGGGAAGAACACCGCTTCA	10	
Db	961	CGCATGAGCGTGTAGACCGGTGCGCAACCGGGGGGAGACGAT	TTGGGAAGAACACCGCTTCA	10	
QY	1021	AGCGCCTGGGAGACCGGGCAACCGTTATCGAATCCTT	CGGAGCCGCGTATTGCGGGCTTC	100	
Db	1021	AGCGCCTGGGAGACCGGGCAACCGTTATCGAATCCTT	CGGAGCCGCGTATTGCGGGCTTC	100	
QY	1081	GCGCGTGAAGTATCTACAAAGAAACCGGACGGACACCCGAC	CAACGAAGATTGAGGGT	114	
Db	1081	GCGCGTGAAGTATCTACAAAGAAACCGGACGGACACCCGAC	CAACGAAGATTGAGGGT	114	
QY	1141	TTCCGCATTCAAGCGCGAACCCGATACGGTCCGGCCGGT	GAGCTTGATTGGCGACCGATC	120	
Db	1141	TTCCGCATTCAAGCGCGAACCCGATACGGTCCGGCCGGT	GAGCTTGATTGGCGACCGATC	120	
QY	1201	ATCGAGCCCGCTGAGTGTATGAGCTTCAGGGGTGGT	TGGAGCGCAGGGGCGCGGCAG	126	
Db	1201	ATCGAGCCCGCTGAGTGTATGAGCTTCAGGGGTGGT	TGGAGCGCAGGGGCGCGGCAG	126	
QY	1261	GGGCTTTC	CGGGGGGCAAGCCATTCTGTCCGCATGAGCAAGCTGTACT	CGAGTGTGGC	132
Db	1261	GGGCTTTC	CGGGGGGCAAGCCATTCTGTCCGCATGAGCAAGCTGTACT	CGAGTGTGGC	132
QY	1321	GCCCTCATGTACTTCGAAAGGCGGGGGAAGAAATGAT	CAAGAGACTTTAACCGCTGCCGTGC	138	
Db	1321	GCCCTCATGTACTTCGAAAGGCGGGGGAAGAAATGAT	CAAGAGACTTTAACCGCTGCCGTGC	138	
QY	1381	CGAAGGTGGTGCAGCCGTCGACACTGAGGACACGAA	GGCAGAGCTAGTCAACGTACACATG	144	
Db	1381	CGAAGGTGGTGCAGCCGTCGACACTGAGGACACGAA	GGCAGAGCTAGTCAACGTACACATG	144	
QY	1441	CGGAGCTGCACACCCGTCGACACTGAGGACACGAA	GGCAGAGCTAGTCAACGTACACATG	144	
Db	1441	CGGAGCTGCACACCCGTCGACACTGAGGACACGAA	GGCAGAGCTAGTCAACGTACACATG	144	
QY	1501	GACGAGAGACGTTGGCGTCTGTGGGAACCCCGCAG	CGACTTGGGCAAGCTCAACTGAG	156	
Db	1501	GACGAGAGACGTTGGCGTCTGTGGGAACCCCGCAG	CGACTTGGGCAAGCTCAACTGAG	156	
QY	1561	GGCGCTGCAAGAGGGGGGGAACGGGCGACCTTGT	TGCGAGCGCGCCGACCGCTTGAA	162	
Db	1561	GGCGCTGCAAGAGGGGGGGAACGGGCGACCTTGT	TGCGAGCGCGCCGACCGCTTGAA	162	
QY	1621	GCGCTTGAAGAGCTGTACGAAGACCGCGCGCAGCG	CGACGAGCCGCTGTGGCAGG	168	
Db	1621	GCGCTTGAAGAGCTGTACGAAGACCGCGCGCAGCG	CGACGAGCCGCTGTGGCAGG	168	
QY	1681	AAGCACTTCCGGAAGCAACAGGACGCGTGTACGCT	CCGGCAGCAAGGGGCGGAAGAGCGG	174	
Db	1681	AAGCACTTCCGGAAGCAACAGGACGCGTGTACGCT	CCGGCAGCAAGGGGCGGAAGAGCGG	174	
QY	1741	CTTGGCGAATCTGAAAGCGCGCGCAAGCCCGAAG	CTTCCCTTGACCATGTGTTCCCGAA	180	
Db	1741	CTTGGCGAATCTGAAAGCGCGCGCAAGCCCGAAG	CTTCCCTTGACCATGTGTTCCCGAA	180	
QY	1801	GACGCGACGTCGACCGCACCGGCCCTTAGTGTGTG	GGGGCGGGTCACTGTAGACAGAC	186	
Db	1801	GACGCGACGTCGACCGCACCGGCCCTTAGTGTGTG	GGGGCGGGTCACTGTAGACAGAC	186	
QY	1861	AAGCGCTGTTGTCGCGGGGCTCTTCGTAAGACAG	ATCTGTTGCACGAAGTGACATCGGCG	192	
Db	1861	AAGCGCTGTTGTCGCGGGGCTCTTCGTAAGACAG	ATCTGTTGCACGAAGTGACATCGGCG	192	
QY	1921	AGGGGCGAGGGAAGCGCCATCGAAGAGCGCGCT	TGTGATCAGTGGGGGAAGCGCGAC	198	
Db	1921	AGGGGCGAGGGAAGCGCCATCGAAGAGCGCGCT	TGTGATCAGTGGGGGAAGCGCGAC	198	
QY	1981	GAGGAGGAGCAACGACGCCACGAGCGGCAACG	GAAGACTGAGCGCGTATG	2031	









693 GCGGAATTGGCGGCTACGTCGGCGGGAAGCCCTTACGGCTTCAGCTTGTTCGGA 752  
 Db 735 GCGGAATTGGCGGCTACGTCGGCGGGAAGCCCTTACGGCTTCAGCTTGTTCGGA 794  
 QY 753 GACGAAGAGATCAGCGGCAACGCGCAATGTGTCATGTCATCAACAAGCTTGGCA 812  
 Db 795 GACGAAGAGATCAGCGGCAACGCGCAATGTGTCATGTCATCAACAAGCTTGGCA 854  
 QY 813 CTCACCACTCCCTTACCGGACCCCTTCGAGTTTCGAGCCGCAAGTAAATCCGCTGTCG 872  
 Db 855 CTCACCACTCCCTTACCGGACCCCTTCGAGTTTCGAGCCGCAAGTAAATCCGCTGTCG 914  
 QY 873 GCGGAGATCAAGAGACAAACCTTCCTTCACAGCCGGGAGTCAAGCCCATTCGA 932  
 Db 915 GCGGAGATCAAGAGACAAACCTTCCTTCACAGCCGGGAGTCAAGCCCATTCGA 974  
 QY 933 CCGGAGGAGATCAGCGGCTTGTATAGCGCATGAGCGTTCGAGCCGTCGACCCGGG 992  
 Db 975 CCGGAGGAGATCAGCGGCTTGTATAGCGCATGAGCGTTCGAGCCGTCGACCCGGG 1034  
 QY 993 CGAGACGATTTGGGAAGAAGACCGCTTCAGCGCTTGGGACCCGGCAACCGTTATGCGAAT 1052  
 Db 1035 CGAGACGATTTGGGAAGAAGACCGCTTCAGCGCTTGGGACCCGGCAACCGTTATGCGAAT 1094  
 QY 1053 CCTTCGGGACCCCGTATTGCGGGCTTCGCGCGCTGAGGATGATCAAGAAGAGCCGGA 1112  
 Db 1095 CCTTCGGGACCCCGTATTGCGGGCTTCGCGCGCTGAGGATGATCAAGAAGAGCCGGA 1154  
 QY 1113 CCGGACCGCGACCAAGATTGAGGTTACCGCATTCAGCGCGACCCGATCAGCTCCG 1172  
 Db 1155 CCGGACCGCGACCAAGATTGAGGTTACCGCATTCAGCGCGACCCGATCAGCTCCG 1214  
 QY 1173 GCGGATGAGCTGATTTGGGAGACCGATTCATCGAGCCCGTGTAGTGTATGACCTTCAGG 1232  
 Db 1215 GCGGATGAGCTGATTTGGGAGACCGATTCATCGAGCCCGTGTAGTGTATGACCTTCAGG 1274  
 QY 1233 GTGGTTGAGCGGAGGCGGCGCGCAAGGGGCTTCCCGGGGCAAGCCATTCGTCGCG 1292  
 Db 1275 GTGGTTGAGCGGAGGCGGCGCGCAAGGGGCTTCCCGGGGCAAGCCATTCGTCGCG 1334  
 QY 1293 CATGGAAGAGCTGTACTGCGAGTGTGCGCGCTCATGACTTGAAGCCGGGGAAGAAATC 1352  
 Db 1335 CATGGAAGAGCTGTACTGCGAGTGTGCGCGCTCATGACTTGAAGCCGGGGAAGAAATC 1394  
 QY 1353 GATCAAGAGCTGTACTGCGAGTGTGCGCGCGCAAGGGTGTGCGACCCGTCGCACTGGGCA 1412  
 Db 1395 GATCAAGAGCTGTACTGCGAGTGTGCGCGCGCAAGGGTGTGCGACCCGTCGCACTGGGCA 1454  
 QY 1413 GCACGAAGGACGTGCACAGTCAGCATGAGCGGCACTCGCAAGTTGTTGCGGAACGAT 1472  
 Db 1455 GCACGAAGGACGTGCACAGTCAGCATGAGCGGCACTCGCAAGTTGTTGCGGAACGAT 1514  
 QY 1473 CTTCAACAAGATCAGGACGCGGAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1532  
 Db 1515 CTTCAACAAGATCAGGACGCGGAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1574  
 QY 1533 GCGCGGAGCTTGGCAAGCTCAGTGAAGCGGCTGAGAAAGCGGCGGAAGAACT 1592  
 Db 1575 GCGCGGAGCTTGGCAAGCTCAGTGAAGCGGCTGAGAAAGCGGCGGAAGAACT 1634  
 QY 1593 TGTTCGGAGAGGCGGCGGAGCGGCTGAGAGCGCTTGAAGAGCTGTGAAGAGAGAGAGAGAG 1652  
 Db 1635 TGTTCGGAGAGGCGGCGGAGCGGCTGAGAGCGCTTGAAGAGAGAGAGAGAGAGAGAGAG 1694  
 QY 1653 AGGCGGATCAGGAG 1712  
 Db 1695 AGGCGGATCAGGAG 1754  
 QY 1713 GCTCCGGAG 1772  
 Db 1755 GCTCCGGAG 1814  
 QY 1773 GCTTCCTTTCACCAATGTTTCCCGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1832

Db 1815 GCTTCCTTTCACCAATGTTTCCCGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1874  
 QY 1833 GTGGTGGGCGCGCGCTCAGTAGACAGACAGCGGCTTGTGTCGAGAGAGAGAGAGAGAGAGAG 1892  
 Db 1875 GTGGTGGGCGCGCGCTCAGTAGACAGACAGCGGCTTGTGTCGAGAGAGAGAGAGAGAGAGAG 1934  
 QY 1893 GATCGTTTCAGGAAGTGCATACGAGGCGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1952  
 Db 1935 GATCGTTTCAGGAAGTGCATACGAGGCGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1994  
 QY 1953 TTTCGATCAGGTGGGCGAAGCCCGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2012  
 Db 1995 TTTCGATCAGGTGGGCGAAGCCCGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2054  
 QY 2013 GGAAGACCTAGCGGCGGTAG 2031  
 Db 2055 GGAAGACCTAGCGGCGGTAG 2073

RESULT 5  
 AX114861  
 LOCUS AX114861  
 DEFINITION Sequence 21 from Patent WO0129208.  
 ACCESSION AX114861  
 VERSION AX114861.1 GI:14031803  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 synthetic construct.  
 artificial sequence.  
 1 (bases 1 to 5711)  
 REFERENCES  
 Kuehn, R., von Melchner, H. and Altschmied, J.  
 Conditional gene trapping construct for the disruption of genes  
 Patent: WO 0129208-A-21 26-APR-2001.  
 JOURNAL  
 ARTEMIS Pharmaceuticals GmbH (DE) ; Frankgen Biotechnologie AG (DE)  
 FEATURES  
 Location/Qualifiers  
 source  
 1..5711  
 /organism="synthetic construct"  
 /db\_xref="taxon:32630"  
 /note="vector pRK65"  
 BASE COUNT 1329 a 1511 c 1617 g 1254 t  
 ORIGIN

Query Match 89.24; Score 1812.6; DB 6; Length 5711;  
 Best Local Similarity 99.88; Pred. No. 0;  
 Matches 1815; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 213 GATGTGAGTACCGCGGCTGTTACGACGTCAGTCCGCGAGGCGAGAAATTCGAGCGC 272  
 Db 1001 GGTGACACGTCACCGGCTGTTACGACGTCAGTCCGCGAGCGAGAAATTCGAGCGC 1060  
 QY 273 AGCAAGCCAGCGACAGAGCGTAGCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 332  
 Db 1061 AGCAAGCCAGCGACAGAGCGTAGCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1120  
 QY 333 AGTTCAGCGGAGAGGCGGCGGTTCAAGTTGTCGCGGCAATTCAGGCAAGCGCGGAGC 392  
 Db 1121 AGTTCAGCGGAGAGGCGGCGGTTCAAGTTGTCGCGGCAATTCAGGCAAGCGCGGAGC 1180  
 QY 393 GTCGCGCTTGGGAGGCGGAGCGCGGAGTTCGAAAGCATCTGAAGAAATTCGCGGCGC 452  
 Db 1181 GTCGCGCTTGGGAGGCGGAGCGCGGAGTTCGAAAGCATCTGAAGAAATTCGCGGCGC 1240  
 QY 453 CCGGCGGCTCAACATGATATGTCATATGACGTCGCGCTTCGCGCTGAGAGTCAAT 512  
 Db 1241 CCGGCGGCTCAACATGATATGTCATATGACGTCGCGCTTCGCGCTGAGAGTCAAT 1300  
 QY 513 GGAAGGATTCGATGTCGGAATTCGCGCTTCGCGCTGAGAGATTCGACTCA 572  
 Db 1301 GGAAGGATTCGATGTCGGAATTCGCGCTTCGCGCTGAGAGATTCGACTCA 1360  
 QY 573 GGAAGGCTTCGCGGAGAGAGAGTGCATGAGACGTCGATTCGATATGAGGCTGCA 632

Db 1361 GGAAGCGCTCTCCGCGCAGGGAACGTCATGACCTGATTCAGCTGATTATGCGCGTCA 1420  
 Qy 633 CGCGTCGACAAAGAAATTTGGTGAAGTGGCGCAAGATTTCTGACACAGAAAGAACTTCA 692  
 Db 1421 CGCCTCGACAAAGAAATTTGGTGAAGTGGCGCAAGATTTCTGACACAGAAAGAACTTCA 1480  
 Qy 693 GCGCGAATTTGGCGGGGACTGTCGCGGGAAGGCGCTTACGCGCTTCAGCTTTGTTTGA 752  
 Db 1481 GCGCGAATTTGGCGGGGACTGTCGCGGGAAGGCGCGCTTACGCGCTTCAGCTTTGTTTGA 1540  
 Qy 753 GACGAGAGATCAACGCGCAACGCGCAAGTGTCAATGTCTCATCAACAAGCTTGGCA 812  
 Db 1541 GACGAGAGATCAACGCGCAACGCGCAAGTGTCAATGTCTCATCAACAAGCTTGGCA 1600  
 Qy 813 CTGACACCTCCCTTACCGGACCTTGCATTTGAGCGCCGCAATACCCGCTGGTGTG 872  
 Db 1601 CTGACACCTCCCTTACCGGACCTTGCATTTGAGCGCCGCAATACCCGCTGGTGTG 1660  
 Qy 873 GCGTGAATCAAGACGCAACAAACCTTCCCTCAAGCGCGGAGTCAAGCGCGCATTC 932  
 Db 1661 GCGTGAATCAAGACGCAACAAACCTTCCCTCAAGCGCGGAGTCAAGCGCGCATTC 1720  
 Qy 933 CCGCGGACATCAACGCGGCTTTGTAGCGCATGAGCGCTGACCGCTCCGACCGCGG 992  
 Db 1721 CCGCGGACATCAACGCGGCTTTGTAGCGCATGAGCGCTGACCGCTCCGACCGCGG 1780  
 Qy 993 CGAGACGATTTGGGAAGACCGCTTCAAGCGCGCTGAGCGCGGACCGCTTATGCGAAT 1052  
 Db 1781 CGAGACGATTTGGGAAGACCGCTTCAAGCGCGCTGAGCGCGGACCGCTTATGCGAAT 1840  
 Qy 1053 CCTTTCGGAACCGCGCTTATGCGCGCTTCCGCTGAGGTGATCTCAAGAAGAACCGCG 1112  
 Db 1841 CCTTTCGGAACCGCGCTTATGCGCGCTTCCGCTGAGGTGATCTCAAGAAGAACCGCG 1900  
 Qy 1113 CGGACCGCGCAACGAGATTTGAGGGTATCCGCAATTAGCGCGACCGCGATCAGCTCG 1172  
 Db 1901 CGGACCGCGCAACGAGATTTGAGGGTATCCGCAATTAGCGCGACCGCGATCAGCTCG 1960  
 Qy 1173 GCGGTCGAGCTTATGTCGCGACGATCATGAGCGCGCTGAGGTGATGAGCTTCAAGC 1232  
 Db 1961 GCGGTCGAGCTTATGTCGCGACGATCATGAGCGCGCTGAGGTGATGAGCTTCAAGC 2020  
 Qy 1233 GTGCTTGGACGCGAGGGGCGCGCAAGGGGCTTTCGCGGGGCGCAAGCATTTCTGCGC 1292  
 Db 2021 GTGCTTGGACGCGAGGGGCGCGCAAGGGGCTTTCGCGGGGCGCAAGCATTTCTGCGC 2080  
 Qy 1293 CATGACAGCTTACTGCGAGTGTGCGCGCTCATGCTTCAAGCGCGGGGGAAGAATC 1352  
 Db 2081 CATGACAGCTTACTGCGAGTGTGCGCGCTCATGCTTCAAGCGCGGGGGAAGAATC 2140  
 Qy 1353 GATCAAGAGCTTACTGCGAGTGTGCGCGAGGGTGTGCAACCGCGTCCGACCTGGGCA 1412  
 Db 2141 GATCAAGAGCTTACTGCGAGTGTGCGCGAGGGTGTGCAACCGCGTCCGACCTGGGCA 2200  
 Qy 1413 GCACGAAGGACGTCAGACGTCAGATGCGCGACTCGACAAGTTCGTTGGGGAACGAT 1472  
 Db 2201 GCACGAAGGACGTCAGACGTCAGATGCGCGACTCGACAAGTTCGTTGGGGAACGAT 2260  
 Qy 1473 CTTCAACAAGTCAAGGCGCGCAAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1532  
 Db 2261 CTTCAACAAGTCAAGGCGCGCAAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2320  
 Qy 1533 CGCCGAGAGCTTGGGCAAGTCACTGAGAGGCGCTGAGAAAGAGGCGGCAAGCTT 1592  
 Db 2321 CGCCGAGAGCTTGGGCAAGTCACTGAGAGGCGCTGAGAAAGAGGCGGCAAGCTT 2380  
 Qy 1593 TGTTCGAGAGCGCGCGAGCGCTTGAAGAGCTGTGAGAGAGCGCGCGCG 1652  
 Db 2381 TGTTCGAGAGCGCGCGAGCGCTTGAAGAGCTGTGAGAGAGCGCGCGCG 2440  
 Qy 1653 AGGCGGCTAGACGAGCGCGCTTGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1712  
 Db 2441 AGGCGGCTAGACGAGCGCGCTTGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2500

Qy 1713 GCTCCGAGAGCAAGGGGCGGAGAGAGCGCTTCCGAACTTGAAGGCGCGAGAGAGAGAG 1772  
 Db 2501 GCTCCGAGAGCAAGGGGCGGAGAGAGCGCTTCCGAACTTGAAGGCGCGAGAGAGAG 2560  
 Qy 1773 GCTTCCCTTGAACAATGTTTCCCGAAGAGACCGAGCGTGAACCGGACCGGCTTAAGTC 1832  
 Db 2561 GCTTCCCTTGAACAATGTTTCCCGAAGAGACCGAGCGTGAACCGGACCGGCTTAAGTC 2620  
 Qy 1833 GTGTGGGGGCGCGCTCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1892  
 Db 2621 GTGTGGGGGCGCGCTCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2680  
 Qy 1893 GATCGTTCACAGAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1952  
 Db 2681 GATCGTTCACAGAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2740  
 Qy 1953 TTGATCAAGTGGGCGAGAGCGCGAGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2012  
 Db 2741 TTGATCAAGTGGGCGAGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2800  
 Qy 2013 GGAAGACGTAGCGCGCTAG 2031  
 Db 2801 GGAAGACGTAGCGCGCTAG 2819  
 RESULT 6  
 PCL414670  
 LOCUS PCL414670 5715 bp DNA circular SYN 19-DEC-2001  
 DEFINITION Cloning vector pSET152.  
 ACCESSION AJ1414670  
 VERSION AJ1414670.1 GI:17974209  
 KEYWORDS lac(3')IV gene; apramycin resistance gene; int gene; integrase; lacZ  
 alpha peptide; lacZ gene.  
 SOURCE Cloning vector pSET152.  
 ORGANISM artificial sequences; vectors.  
 REFERENCE 1  
 AUTHORS Wilkinson, C.J., Hughes-Thomas, Z.A., Martin (nee Rowe), C.J.,  
 Bohm, I., Mironenko, T., Deacon, M., Wheatcroft, M., Wirtz, G.,  
 Staunton, J. and Leadlay, P.F.  
 TITLE Increasing the efficiency of heterologous promoters in  
 actinomycetes  
 JOURNAL J. Mol. Microbiol. Biotechnol.  
 REFERENCE 2 (bases 1 to 5715)  
 AUTHORS Wilkinson, C.J.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-SEP-2001) Wilkinson C.J., Department of Biochemistry,  
 University of Cambridge, 80 Tennis Court Rd., Cambridge, CB2 1GA,  
 UNITED KINGDOM  
 COMMENT Related sequences V00635, J02451, M24524, M21778.  
 FEATURES  
 source  
 1..5715  
 /organism="Cloning vector pSET152"  
 /specific\_host="Streptomyces sp."  
 /db\_xref="taxon:173260"  
 /lab\_host="Escherichia coli"  
 complement(163..507)  
 /gene="lacZa"  
 complement(163..507)  
 /gene="lacZa"  
 complement(163..507)  
 /function="blue-white screening"  
 /codon\_start=1  
 /transl\_table=11  
 /product="lacZ alpha peptide"  
 /protein\_id="CAC93946.1"  
 /db\_xref="GI:17974210"  
 translation="MTRNSISRAADPLETSCPSLAVALVLRDMDENPGVTOURL  
 AAHPFASMRNSEEARTRPSQQLRSLNGEWRMRVFLTLGCSHRIINPMSTSG  
 IGSAADAKRST"  
 complement(508..509)  
 /note="pUC18"  
 /replac="ggt"  
 variation

gene 1729..2505  
/gene="aac(3)IV"  
1729..2505  
/gene="aac(3)IV"  
/function="apramycin resistance"  
/codon\_start=1  
/transl\_table=11  
/product="apramycin resistance gene"  
/protein\_id="CAC93947.1"  
/db\_xref="GI:17974211"  
/translation="MOYEMRKAEILIGOLNLGYTPGGVILLVHSSPFRVRLNDEPIGL  
IEALRALGPGGLVMPMSGLDEPDPTSPYDPLVSTPTKRLPVKSAHP  
APAAQPOAEQIISDPLRPHSPASVAVVHELDQVLLGSHDANTLHAEIEMA  
KVPYGPBRTCTIQQDKLVVDLENDCCERALLADRWLEKSLQKEGPHAFARL  
IKSRDIVATLGGDRLFLHPPEAGCEECDAARQSLG"  
340..3451  
/note="R4 origin of single-stranded DNA transfer"  
/direction=RIGHT  
3799..3837  
/note="phic31 attP site"  
3854..5695  
/gene="int"  
3854..5695  
/gene="int"  
/function="phic31 integrase"  
/codon\_start=1  
/transl\_table=11  
/product="Integrase"  
/protein\_id="CAC93948.1"  
/db\_xref="GI:17974212"  
/translation="MTQGVVTVGVDYVAGAYDROSPRENSAASPATORSANEDKRAAD  
LOREVRDGRFREFVGFSEAPGTSAFGEABERPEPERILNECRAGLNNTIYYDSRF  
SRKLVMDAIPIYSLEIALLGVTVSTQEGFPROGNWMDLHLIMRLDASKESSLASAK  
IIDTKLORELGIVYVGKAPYFELVSEKTEITRGRMNVAIVINKLAHSTPLTGFE  
FEBDVIKRMREIKHKHLPKPGSOALHPGSLITGLCKRMDADVPTGETIGKTA  
SSAMDAIVMRLIRDFRIAGFAEVIYKPKPDPTTKLEGRIORDPLTLPEVLD  
GPIIEPAEYVELQALDGRGKGLSRQALISAMDKLYCEGAVTSKGESEISKS  
YRCRRKVVDPAPGOHECTCNVSMALDKFAVERFINKIRHAEDEETLALIMEAR  
RFGKLEAPEKSGERANLVABRADALNLEELTEPRACAVYGPGRKFRQQAALT  
LRQGAERLAELEAEAPKPLIDOMFPEDADADPTGPKSMRNGASVDDKRVFGLV  
DKIVYTKSTTGCGCPITKRASTIMAKPTDDDEDADGDIEDVAA"

BASE COUNT 1189 a 1643 c 1745 g 1138 t  
ORIGIN

Query Match 89.2%; Score 1812.6; DB 12; Length 5715;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1815; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 213 GATGTGAGGTACGCGGCTGCTTACGACCGTACGTGCGCGAGCGGAGAAATTCGAGCGC 272  
DB 3877 GGTGGACAGCTACGCGGCTGCTTACGACCGTACGTGCGCGAGCGGAGAAATTCGAGCGC 3936  
QY 273 AGCAAGCCAGCAGCAGCAGGTAGCGCAAGCAAGAGAGGCGCGGCGGCGGCGGCGCA 332  
DB 3937 AGCAAGCCAGCAGCAGCAGGTAGCGCAAGCAAGAGAGGCGCGGCGGCGGCGGCGCA 3996  
QY 333 AGTCAGCGCAGCAGCGGCGGCTTACGTTCTGTCGGGCAATTCAGGAGAGCGCGGCGC 392  
DB 3997 AGTCAGCGCAGCAGCGGCGGCTTACGTTCTGTCGGGCAATTCAGGAGAGCGCGGCGC 4056  
QY 393 GTGCGGCTTGGGAGCGGCGGAGCGCGGAGTTTCGAGACGATCTTAAGCAATTCGCGCGC 452  
DB 4057 GTGCGGCTTGGGAGCGGCGGAGCGCGGAGTTTCGAGACGATCTTAAGCAATTCGCGCGC 4116  
QY 453 CGGCGGCTTCAACATGATGATGTCATGAGTGTGCGGCTTCTGCGGCTTAAGAGTCAAT 512  
DB 4117 CGGCGGCTTCAACATGATGATGTCATGAGTGTGCGGCTTCTGCGGCTTAAGAGTCAAT 4176  
QY 513 GACAGCGATTCGATGATGTCGGAATTCGCGCTTGGGCGTGAAGATTCGTTCCACTCA 572  
DB 4177 GACAGCGATTCGATGATGTCGGAATTCGCGCTTGGGCGTGAAGATTCGTTCCACTCA 4236  
QY 573 GGAAGGCGCTTCGCGGAGGAGAAAGCTATGAGCTGATTCAGCTGATTAAGCGGCTCGA 632

DB 4237 GGAAGGCGCTTCGCGGAGGAGAAAGCTATGAGCTGATTCAGCTGATTAAGCGGCTCGA 4296  
QY 633 CGGCTCCACAAAGATCTTCGCTGAAGTCGGGAGAGATTCGACACGAGAACCTTCA 692  
DB 4297 CGGCTCCACAAAGATCTTCGCTGAAGTCGGGAGAGATTCGACACGAGAACCTTCA 4396  
QY 693 GCGGAAATTTGGGCGGAGCTGCGGCGGAGGCGGCTTACGGCTTCGAGCTTGTTCGGA 752  
DB 4357 GCGGAAATTTGGGCGGAGCTGCGGCGGAGGCGGCTTACGGCTTGTTCGGA 4416  
QY 753 GACGAAGAGATCAACGGCAACGCGCAATGTCATGTCATCAACAGCTTGGCGCA 812  
DB 4417 GACGAAGAGATCAACGGCAACGCGCAATGTCATGTCATCAACAGCTTGGCGCA 4476  
QY 813 CTGACACACTCCCTTACCGGAGCCCTTCGAGTTCGAGCCGACGTAATCCGTGTGTG 872  
DB 4477 CTGACACACTCCCTTACCGGAGCCCTTCGAGTTCGAGCCGACGTAATCCGTGTGTG 4536  
QY 873 GCGTAGATCAAGACGACACACACCTCCCTCAAGCGGCGAGTCAAGCGGCAATTC 932  
DB 4537 GCGTAGATCAAGACGACACACACCTCCCTCAAGCGGCGAGTCAAGCGGCAATTC 4596  
QY 933 CCGGCGAGCATACGAGGCTTGTAAAGCGCATGACGCTGACGCGGTCCGACCGCGG 992  
DB 4597 CCGGCGAGCATACGAGGCTTGTAAAGCGCATGACGCTGACGCGGTCCGACCGCGG 4656  
QY 993 CGAGAGATTTGGGAGAGAGACCGCTTCAAGCGGCTGGAGACCGGCAACCGTTATGGAAT 1052  
DB 4657 CGAGAGATTTGGGAGAGAGACCGCTTCAAGCGGCTGGAGACCGGCAACCGTTATGGAAT 4716  
QY 1053 CCTTGGGAGCCCGGTATTGCGGGCTTCGCGCTGAGGTATCTACAAAGAGCCGGA 1112  
DB 4717 CCTTGGGAGCCCGGTATTGCGGGCTTCGCGCTGAGGTATCTACAAAGAGCCGGA 4776  
QY 1113 CGGCGAGCGGACAGCAACATTTAGAGGTATACCGCATTCAGCGGAGCCGATACGCTCG 1172  
DB 4777 CGGCGAGCGGACAGCAACATTTAGAGGTATACCGCATTCAGCGGAGCCGATACGCTCG 4836  
QY 1173 GCGGCTGAGCTGATTTGGCGAGCCGATCATGAGCCCGCTGATGATGATGATGATGATG 1232  
DB 4837 GCGGCTGAGCTGATTTGGCGAGCCGATCATGAGCCCGCTGATGATGATGATGATGATG 4896  
QY 1233 GTGCTTGGAGCGGAGGCGGCGCGGCAAGGCGCTTCCGCGGCGCAAGCCATTCGTCGC 1292  
DB 4897 GTGCTTGGAGCGGAGGCGGCGCGGCAAGGCGCTTCCGCGGCGCAAGCCATTCGTCGC 4956  
QY 1293 CATGCAAGCTGTACTGAGTGTGCGCGCGTCAATGATTTGGAAGCGGAGGAAGATTC 1352  
DB 4957 CATGCAAGCTGTACTGAGTGTGCGCGCGTCAATGATTTGGAAGCGGAGGAAGATTC 5016  
QY 1353 GATCAAGGACTTACCGCTGCGTGCCTGCGGAGAGGTGTGCGACCGCTCCACTCGGCGCA 1412  
DB 5017 GATCAAGGACTTACCGCTGCGTGCCTGCGGAGAGGTGTGCGACCGCTCCACTCGGCGCA 5076  
QY 1413 GCAGCAAGGCAAGTCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1472  
DB 5077 GCAGCAAGGCAAGTCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 5136  
QY 1473 CTTCAACAAGATCAAGGCAAGGCGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1532  
DB 5137 CTTCAACAAGATCAAGGCAAGGCGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5196  
QY 1533 CGCGCGAGCTTGGGCAAGCTCACTGAGGCGCTTGAAGAGAGCGGCGAAGCGGCAACCT 1592  
DB 5197 CGCGCGAGCTTGGGCAAGCTCACTGAGGCGCTTGAAGAGAGCGGCGAAGCGGCAACCT 5256  
QY 1593 TGTTCGAGAGGCGGCGGAGCGGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1652  
DB 5257 TGTTCGAGAGGCGGCGGAGCGGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5316  
QY 1653 AGGCGCTTACGAG 1712

Db 5317 AGCGCCGACGACGAGCCGTTGGCAGAGACCTTCGGAAGCAACAGCAGCGCTGAC 5376  
Oy 1713 GCTCCGAGCAAGAGGGGGGAGAGAGCGGCTTCGCAACTTGAAGCGCCGAGCCCGCA 1772  
Db 5377 GCTCCGAGCAAGAGGGGGGAGAGAGCGGCTTCGCAACTTGAAGCGCCGAGCCCGCA 5436  
Oy 1773 GCTTCCCTTGACCAATGTTCCCGGAAGACGCGCAGCTGACCCGAGCCGCTTAAGTC 1832  
Db 5437 GCTTCCCTTGACCAATGTTCCCGGAAGACGCGCAGCTGACCCGAGCCGCTTAAGTC 5496  
Oy 1833 GTGGTGGGGGGGGCGGCTGACGACGACGAGCGGCTGTTGCTGGGGCTCTTGACACAA 1892  
Db 5497 GTGGTGGGGGGGGCGGCTGACGACGAGCAAGCGGCTGTTGCTGGGGCTCTTGACACAA 5556  
Oy 1893 GATCGTTGTACGAGTGCAGTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1952  
Db 5557 GATCGTTGTACGAGTGCAGTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5616  
Oy 1953 TTGCATCACGTTGGGGGAGAGCCGCGACGACGACGACGACGACGACGACGACGACGAC 2012  
Db 5617 TTGCATCACGTTGGGGGAGAGCCGCGACGACGACGACGACGACGACGACGACGACGAC 5676  
Oy 2013 GGAAGACGTAGCGGGCTAG 2031  
Db 5677 GGAAGACGTAGCGGGCTAG 5695

RESULT 7  
BPH6589  
LOCUS BPH6589 41489 bp DNA linear PHG 12-MAY-2000  
DEFINITION Bacteriophage phi-C31 complete genome.  
ACCESSION AJ006589  
VERSION AJ006589.1 GI:3947449  
KEYWORDS C gene: dCMP deaminase; deoxynucleotide monophosphate kinase; DNA polymerase; int gene; integrase; large subunit; major capsid protein; major tail protein; portal protein; protease; putative; repressor; tail fibre protein; tail tape measure protein; terminase; transfer RNA-Thr; tRNA-Thr.  
Bacteriophage phi-C31.  
Bacteriophage phi-C31.  
Virus; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae; Lambda-like viruses.  
1 (bases 1 to 41489)  
Hendrix, R.W., Smith, M.C., Burns, R.N., Ford, M.E. and Hatfull, G.F.  
Evolutionary relationships among diverse bacteriophages and prophages: all the worlds a phage  
Proc. Natl. Acad. Sci. U.S.A. 96 (5), 2192-2197 (1999)

SOURCE ORGANISM  
JOURNAL MEDLINE 99162580  
PUBMED 10051617  
2 (bases 1 to 41489)  
Smith, M.C., Burns, R.N., Wilson, S.E. and Gregory, M.A.  
The complete genome sequence of the Streptomyces temperate phage straight phiC31: evolutionary relationships to other viruses  
Nucleic Acids Res. 27 (10), 2145-2155 (1999)  
JOURNAL MEDLINE 99238410  
PUBMED 10219087  
3 (bases 1 to 41489)  
Smith, M.C.M.  
Direct submission  
Submitted (01-JUN-1998) Smith M.C.M., Genetics, University of Nottingham, Queens Medical Centre, Nottingham, NG7 2UH, UK  
FEATURES  
source  
1. 41489  
/organism="Bacteriophage phi-C31"  
/vifion  
/strain="Norwich stock"  
/db\_xref="taxon:10719"  
145..585  
/gene="31"  
145..585  
/gene="31"  
/codon\_start=1  
/transl\_table=1

gene  
CDS  
/product="gp31"  
/protein\_id="CA07101.1"  
/db\_xref="GI:3947450"  
/db\_xref="SPTREMBL:Q37835"  
/translation="WSRAKSPDLRTGNNANNAEASAPVYEGRAFRPAHLKATGKV  
WRNVAQMGAVSPDTRNVILRYCELDHRRADLLLEADYNSGNGQVPAHPML  
RYESTERELRSIEFAIGFTPEARMRRLGIVAAEARKVAAGPEDF"  
complement(640..1170)  
/gene="32"  
complement(640..1170)  
/note="transcribed in lysogens"  
/codon\_start=1  
/transl\_table=11  
/product="gp32"  
/protein\_id="CA07102.1"  
/db\_xref="GI:3947451"  
/db\_xref="SPTREMBL:Q37837"  
/translation="MNANATPTRYFGKTNGLHYVYVLRGELRYTPGYTLDMNAVAD  
AACWEAFVVAAPAEYPAERDRMRDEAGVADKPVGELREGDYVNHVTYTKAAVNRKG  
ORVSEYOVSEWVECSGAWARDMARGLVGYRLVGFYANKGETVYVHRPRLAKR  
AAEATRAVYAAFSKR"  
1282..2841  
/gene="33"  
1282..2841  
/gene="33"  
/function="putative large terminase subunit"  
/codon\_start=1  
/transl\_table=11  
/product="gp33"  
/protein\_id="CA07103.1"  
/db\_xref="GI:3947452"  
/translation="MADNAGIDPVIARHHPADATVPSRGYRAKWIIEECYLTGSFAG  
OPRLLPQRELLIDAYVLTQDTEGRMRKRHTVYVCARKKSKTIAAAILYLLIA  
DRGDRQRIITAAANDRONRVDSAKQVNNAPRLAACDQROVIRKDDTYRVS  
ADAGQOGLNPAAVSLDEAFSRHSKDLDPALILGSAARQPMFLIISAGPDGFA  
ALCEQGERVNSGEADDPILFYFSWGPILGELVDHPVWRACNSYDILNDDPKAA  
AORSTEAFRIYRLSOFVWGAFTPLPHGLMDSLAADDPLEPGEVIGFDSMRGDS  
TALVACRIKRLKLVFGHWEAPADDAHVRVPAVDVREELHTALDYYRVNLYADPYRW  
EETLDNLNLEADGPVPAFPNLSARVPATQAVYDCCRGRSHSDNPRALGRITGNVIL  
KEDARGARITKEHASSRRKIDLAVMVLAVHGAVMVREDNGIVSDKPIITATWEDDEGN  
VFVHPDHAFF"  
2886..4259  
/gene="34"  
2886..4259  
/gene="34"  
/function="putative portal protein"  
/codon\_start=1  
/transl\_table=11  
/product="gp34"  
/protein\_id="CA07104.1"  
/db\_xref="GI:3947453"  
/db\_xref="SPTREMBL:Q92XB2"  
/translation="MGFWSALFGRGHSPALDIEARAEPPYDPSIYNLGAVASGETV  
TPHDALQVSAVSAVRLSETIATPLSTYSKRGSKREIYVPEWLDYPMNEPGCMGR  
IDILSOTVSLILQGNAPFLAVWOGPNITVGLDVLPTKIHWVWVWDLRKRFVPAVD  
IDADGNVILGCFMTPRQVLIHIDGMILPDDPVCSPISYARSIGLALAAOKSGSEFA  
NGAMGAVYVPGTMSBEGILARAEARANSQVNAHVALITLGAESKAMSPDE  
AOFLOTROFOVPELIRIFGVPHLISDNTNSTWSGLAEQNIATFTMSFLRWLERIE  
AGFNILFAETADRRFPVKFNIDELIKRAPKRMELWSGLONGIYSIDEVAAADMT  
PLPDGLGKRYVPLNIGVGEPEPEPAPAPAIIPPAEPEDEPEPEPEPEPEPEPEPEPE  
EDDEDDA"  
4240..4881  
/gene="35"  
4240..4881  
/gene="35"  
/note="putative protease"  
/codon\_start=1  
/transl\_table=11  
/product="gp35"  
/protein\_id="CA07105.1"  
/db\_xref="GI:3947454"  
/db\_xref="SPTREMBL:Q92XB1"



gene  
CDS  
/translation="MKTRRLREFAVGLEERASDDGRISMGRYAAYFENLSDLGE  
RERIVPAGAPSLQNDVYATFNDHTRSLGRSSGTLRVGEDEEGWYEDIDPDTY  
GRVADKLKRGDLOGSSFTVRIDGGGRADDDDPENGLPVRETTADMDVETLPPVNP  
AYPTTASLRISIEETLRIGEFAPTTERDSQPOGDVAPASHFDRSLVRLAKSK  
4893. 6071  
/gene="36"  
4893. 6071  
/gene="36"  
/function="major capsid protein"  
/codon\_start=1  
/transl\_table=11  
/product="gp36"  
/protein\_id="CAA07106.1"  
/db\_xref="GI:3947455"  
/db\_xref="SPTREMBL:092WV6"  
/translation="MDATTLSEANEERATLRLTDEFAGKEMTAEREKEERLL  
TAAVDFGRIRKIDAIKADAVTSLISGQSGSGAQRASDADDVAVLRAGLGER  
SFEAPKRCGTAKGNPNVLSRLYGLIAQVRSKIMRGASTITSDANEMDTY  
ITGRATAGIVGETAIEPSTPATQNSMGSKFYGVASVEFATDQVLDLGVFLVSD  
AGPAIDGMRHFLTGTGTQPRGILTDAGANAAEGADADSKVDALIDLHEVPS  
AYRNMKFEVNDLRAOMRKIKDANGQYLSALITGAPDTFGKVEVETDGDMPADKY  
LFADLSKYRVRFAAGSLRVRSDAKFSTIDQIVRFIQRADGLLVARCAKVLVTPAA  
"  
mat\_peptide  
5226. 6068  
/gene="36"  
/product="gp36"  
6144. 6512  
/gene="37"  
6144. 6512  
/gene="37"  
/codon\_start=1  
/transl\_table=11  
/product="gp37"  
/protein\_id="CAA07107.1"  
/db\_xref="GI:3947456"  
/db\_xref="SPTREMBL:092XW0"  
/translation="MAYATIEELRALDGDSDSLFEDELISDAIDPSVETVAYCCK  
MDNADPTPEPTIMCVRLARQYVLDHVSRIPOKALQLOSEFGSIOLAQAGNWRPES  
LPEVNKKILNRYRLEPFIEM"  
gene  
CDS  
6519. 6956  
/gene="38"  
6519. 6956  
/gene="38"  
/codon\_start=1  
/transl\_table=11  
/product="gp38"  
/protein\_id="CAA07108.1"  
/db\_xref="GI:3947457"  
/db\_xref="SPTREMBL:092XA9"  
/translation="MALIFNAKVRLEFALKANVPDGYOCTFAETGDSNRKRWIGAT  
VDDDLAPVAMBSGAKPTNVTGYEAAHAYTTGPNPDAERANYGIRDIYKAKCAAVNA  
DLVSYVGLMDVPRBSASVSTETTDGASALTIVRVRGRVY0"  
6589. 7975  
/gene="39"  
6589. 7975  
/gene="39"  
/function="major tail protein"  
/codon\_start=1  
/transl\_table=11  
/product="gp39"  
/protein\_id="CAA07109.1"  
/db\_xref="GI:3947458"  
/db\_xref="SPTREMBL:092XAB"  
/translation="MALDASISIGREDYTGTLISGVEYEGCAQDSWKOTVEIESVGF  
RAGKOTARADRRNMVNGEGELEIDVADAGASLLTAFFKVTYTDGCVKTYVLET  
SDVSAFSTLAQVAPRGVDSKARHLGCATWESLAVEEAVKLAIVTEDEFRLATH  
TSTPAIVEPTPAEYAPYDWTFTAVELSRAGSVYEPATSLIELGELGMDTDRFLN  
GSELKKKVRNALPTYEGTLEGEFSAASIGLIDAVAGEVCSFKATFEFGIILGSSLSLV  
EAPAIQFTGESSEATDEVTIHNLPFRVLRODGDGTPAIKILTYERGTPIVER"  
7975. 8406  
/gene="40"  
7975. 8406  
/gene="40"  
/gene="40"

Query Match	Best Local Similarity	Score	1812.6;	DB 7;	Length	41489;
Matches	1815;	Conservative	0;	Mismatches	4;	Indels
gene						
8406. 8878						
213	GATGTGAGGTACGCGGCTTACAGCGTCAAGTCGCGGAGCGGAGATTGAGCGC	272				
38444	GTTGGACACGTACGCGGCTTACAGCGTCAAGTCGCGGAGCGGAGATTGAGCGC	38503				
273	AGCAAGCCGACGACACAGCGTACGCGGAGCGGAGCGGAGCGGAGCGGAGCGC	332				
38504	AGCAAGCCGACGACACAGCGTACGCGGAGCGGAGCGGAGCGGAGCGGAGCGC	38563				
333	AGTCGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGC	392				
38564	AGTCGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGC	38623				
393	GTCGCGGCTTCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGC	452				
38624	GTCGCGGCTTCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGC	38683				
453	CGGCGGCTTCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGC	512				
38684	CGGCGGCTTCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGC	38743				
513	GGACCGGATTCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGC	572				
38744	GGACCGGATTCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGC	38803				
573	GGAGCGGCTTCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGC	632				
38804	GGAGCGGCTTCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGC	38863				
633	CGCGTGCACAAAGAAATCTTCGCTGAAGTCGCGGAGCGGAGCGGAGCGC	692				
38864	CGCGTGCACAAAGAAATCTTCGCTGAAGTCGCGGAGCGGAGCGGAGCGC	38923				
693	GGCGGATTCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGC	752				
38924	GGCGGATTCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGC	38983				
753	GAGGAGGAGATCACGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGC	812				
38984	GAGGAGGAGATCACGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGC	39043				
813	CTCGACCATCCCTTACCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGC	872				
39044	CTCGACCATCCCTTACCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGC	39103				
873	GGTGAGATCAAGACGACAAACACCTTCCTCAAGCGGAGCGGAGCGGAGCGC	932				
39104	GGTGAGATCAAGACGACAAACACCTTCCTCAAGCGGAGCGGAGCGGAGCGC	39163				
933	CCCGGAGCATCACGCGGCTTGTAAAGCGATGAGACCTTGAGCGGAGCGGAGCGC	992				
39164	CCCGGAGCATCACGCGGCTTGTAAAGCGATGAGACCTTGAGCGGAGCGGAGCGC	39223				
993	CGAGACGATGGGAGGAGACCGCTTCAAGCGGAGCGGAGCGGAGCGGAGCGC	1052				
39224	CGAGACGATGGGAGGAGACCGCTTCAAGCGGAGCGGAGCGGAGCGGAGCGC	39283				
1053	CTTCGGGAGCGGAGCGGCTTGTAAAGCGATGAGACCTTGAGCGGAGCGGAGCGC	1112				



Db 39284 CCTTCGGAGCCCGGCTATTGGCGGCTTCGCCGCTGAGTGATCTACAAGAAAGCCGGA 39343  
QY 1113 CGGACGCGCCGACGAAAGATTGAGGGTTACCGATTAGGGCGACCCGATCACGCTCG 11172  
Db 39344 CGGACGCGCCGACGAAAGATTGAGGGTTACCGATTAGGGCGACCCGATCACGCTCG 39403  
QY 1173 GCCGCTGAGGCTTATTCGCGAGCCGATCATGAGCCCGCTAGTGTATGAGCTTCAGGC 1232  
Db 39404 GCGGCTGAGGCTTATTCGCGAGCCGATCATGAGCCCGCTAGTGTATGAGCTTCAGGC 39463  
QY 1233 GTGGTTGAGCGGACAGGGGGCGCGCAAGGGGCTTTCCGGGGGCAAGCCATTCTGTCCGC 1292  
Db 39464 GTGGTTGAGCGGACAGGGGGCGCGCAAGGGGCTTTCCGGGGGCAAGCCATTCTGTCCGC 39523  
QY 1293 CATGACAAAGCTTACTGCGAGTGCGGGCGCTCATGACTTCGAAAGCCGCGGGGAAGATC 1352  
Db 39524 CATGACAAAGCTTACTGCGAGTGCGGGCGCTCATGACTTCGAAAGCCGCGGGGAAGATC 39583  
QY 1353 GATCAAGGACTCTTACCGCTCCGCTCCGCGGAGAGTGCTGACCCGCTCCGACCTGGGCA 1412  
Db 39584 GATCAAGGACTCTTACCGCTCCGCTCCGCGGAGAGTGCTGACCCGCTCCGACCTGGGCA 39643  
QY 1413 GCACGAAGGACGCTGCAACGTCACGATGGCGGACCTCGACAAGTTGTTGGCGAAGCAT 1472  
Db 39644 GCACGAAGGACGCTGCAACGTCACGATGGCGGACCTCGACAAGTTGTTGGCGAAGCAT 39703  
QY 1473 CTTCACAAAGTACAGGCAAGCCGGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1532  
Db 39704 CTTCACAAAGTACAGGCAAGCCGCGGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 39763  
QY 1533 CGCCGACGCTTCCGCAAGCTCAGTGAAGGCGCTGAGAAAGAGCGGCGCAAGCGCAACT 1592  
Db 39764 CGCCGACGCTTCCGCAAGCTCAGTGAAGGCGCTGAGAAAGAGCGGCGCAAGCGCAACT 39823  
QY 1593 TGTTCGCGAGCGCGCGCGACGCTTGAAGCGCTTGAAGAGCTGTACGAAGACCGCGCGC 1652  
Db 39824 TGTTCGCGAGCGCGCGCGACGCTTGAAGCGCTTGAAGAGCTGTACGAAGACCGCGCGC 39883  
QY 1653 AGGGCGGACGAG 1712  
Db 39884 AGGGCGGACGAG 39943  
QY 1713 GCTCGCGGACGAG 1772  
Db 39944 GCTCGCGGACGAG 40003  
QY 1773 GCTTCGCTTACCAAGTGTTCGCGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1832  
Db 40004 GCTTCGCTTACCAAGTGTTCGCGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 40063  
QY 1833 GTGGGGGGGGGGCGGTAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1892  
Db 40064 GTGGGGGGGGGGCGGTAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 40123  
QY 1893 GATCGTTGTCAGAAAGTCTAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1952  
Db 40124 GATCGTTGTCAGAAAGTCTAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 40183  
QY 1953 TTCGATCACGTTGGGGAAG 2012  
Db 40184 TTCGATCACGTTGGGGAAG 40243  
QY 2013 GGAAGAGCTAGCGGCGTAG 2031  
Db 40244 GGAAGAGCTAGCGGCGTAG 40262

RESULT 8  
EV084006 5967 bp DNA circular SYN 03-MAY-1997  
LOCUS  
DEFINITION Expression vector pBSII-LUCINT firefly luciferase (LUCINT),  
beta-galactosidase (lacZ) and beta-lactamase (ampR) genes, complete  
cds and lac operon promoter sequence.

ACCESSION U84006  
VERSION U84006.1 GI:2071944  
KEYWORDS  
SOURCE Expression vector pBSII-LUCINT.  
ORGANISM Expression vector pBSII-LUCINT  
REFERENCE artificial sequences; vectors.  
AUTHORS 1 (bases 1 to 5967)  
Mankin, S.L., Allen, G.C. and Thompson, W.F.  
TITLE Introduction of a plant intron into the Luciferase Gene of Photinus  
pyralis  
JOURNAL Plant Mol. Biol. Rep. (1997) In press  
REFERENCE 2 (bases 1 to 5967)  
AUTHORS Mankin, S.L., Allen, G.C. and Thompson, W.F.  
TITLE Direct Submission  
JOURNAL Submitted (06-JUN-1997) Botany, NC State Univ., Box 7612, Raleigh,  
NC 27695-7612, USA  
FEATURES  
source  
1..5967  
/organism="Expression vector pBSII-LUCINT"  
/db\_xref="taxon:56405"  
/plasmid="pluk07"  
/note="chimera plasmid consisting of a pBS-type plasmid  
(pluk07, dsDNA) derived from the LUC\* gene (a synthetic  
cdna gene) by the addition of a plant intron sequence"  
52..851  
/note="Cawv 35S promoter"  
875..3052  
/gene="LUCINT"  
join(875..1029,1219..2716)  
/gene="LUCINT"  
/EC\_number="1.13.12.7"  
/note="luciferin 4-monooxygenase; LUC\* with PIV2 intron"  
/codon\_start=1  
/product="firefly luciferase"  
/protein\_id="AAB53627.1"  
/db\_xref="GI:2071945"  
/translation="MEDAKNIKKGPPAPPPLEDGTAGBOLHAKKRRALVPGTIAFTD  
AHIEVITVAEYFEMSVRLAEAMKRYGKINSTRIVVCSSENSIOFPMVIALFTGVAV  
APANDIYNERELLSNMISOPTVYFVSKGQIKTLINOKRIPIIOKITIMSKPDYCG  
FOSMTFVTSHPFNDFNDFPSFDBDKTIALIMSSGSLPKGVALDHRACRA  
FSHARDIPFNQIIPDVALISVPFHGFGFTLLGLICGRVLLWREEELFLS  
LQYRIQISALVPTLPSFFAKSLIDKXDSLNLHEISGAPLSKEVGAARFHLF  
GIRQGYGTEVNTTSAIITPEBDDKRGAVKVPPEFAKVVDLDYGTGLVQSGELV  
RGPIMSGYVNPATNALIDKQGLMSGDIAVYDEDEHFTVRLSLIKYQYOA  
PALESLILOHNPEDAGVACLPPDDGGEIPAAVYVEHCKTMEKEIVDVVASQVTT  
AKKLRGVYFVDEVPKILTKLIDARKIREILIKAKKGSKL"  
1030..1218  
/gene="LUCINT"  
/note="PIV2 chimeric intron"  
2786..3052  
/gene="LUCINT"  
/standard\_name="nopoline synthase 3' region"  
/note="nos"  
2816..2821  
/gene="LUCINT"  
3146..3355  
/gene="lacZ"  
3146..3355  
/gene="lacZ"  
/codon\_start=1  
/transl\_table=11  
/product="beta-galactosidase"  
/protein\_id="AAB53629.1"  
/db\_xref="GI:2071947"  
/translation="MAVYLORDEMENGVTQNLRLAAHPAPASWRNSEARTDRPSOO  
LRSLNGEMREIVNVNILLKFLALNFC"  
3895..4755  
/gene="ampR"  
3895..4755  
/gene="ampR"  
/function="ampicillin resistance marker"  
/codon\_start=1  
/transl\_table=11

[illegible]

QY	201	CCAAATTTGTTGATGTGCAGTACCGGGGTGC	233
Dd	445	CCAAAATTGTTGATGTGCAGTGTCAGAGGCCG	477
RESULT 10			
AX259244			
LOCUS		AX259244	1577 bp DNA linear
DEFINITION		Sequence 42 from Patent WO0173087.	PAT 26-OCT-2001
ACCSSION		AX259244	
VERSION		AX259244.1 GI:16508490	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS		Hohn,T., Stavolone,L., de Haan,P.T., Ligon,H.T. and Kononova,M.	
TITLE		Cestrum yellow leaf curling virus promoters	
JOURNAL		Patent: WO 0173087-A 42 04 OCT-2001;	
FEATURES		Syngenta Participations AG (CH)	
source		Location/Qualifiers	
		1..1577	
		/organism="synthetic construct"	
		/db_xref="taxon:32630"	
		/note="artificial sequence Cmpc-synGPPI-nos expression cassette"	
BASE COUNT		450 a 372 c 390 g 365 t	
ORIGIN			
Query Match		9.8%; Score 198.6; DB 6; Length 1577;	
Best Local Similarity		95.8%; Pred.No. 7.4e-27;	
Matches 204; Conservative		0; Mismatches 9; Indels 0; Gaps 0;	
QY	21	CGGGGAGTAGTACTAGTCTTCGCTTCACTTGATATAATTAATTAATCATTA	80
Dd	645	CCGGAGGGCTAGCTAGTGTTCCTTCACCTTGATATAATTAATTAATTAATTA	704
QY	81	TTAGTAGTAATTAATTAATTTCAAATATTTTTTTCAAAATAAAGAATAGATAT	140
Dd	705	TTAGTAGTAATTAATTAATTTCAAATATTTTTTTCAAATAAAGAATAGATAT	764
QY	141	ATTTGCTTTCTCCTACTTTATATAGTGATATTTTAATTTTACTTTCTATATGA	200
Dd	765	ATTTGCTTTCTCCTACTTTATATAGTGATATTTTAATTTTACTTTCTATATGA	824
QY	201	CCAAAATTGTTGATGTGCAGGTACCGGGGTGC	233
Dd	825	CCAAAATTGTTGATGTGCAGGTACCGGGGTGC	857
RESULT 11			
AX259242			
LOCUS		AX259242	1618 bp DNA linear
DEFINITION		Sequence 40 from Patent WO0173087.	PAT 26-OCT-2001
ACCSSION		AX259242	
VERSION		AX259242.1 GI:16508488	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS		Hohn,T., Stavolone,L., de Haan,P.T., Ligon,H.T. and Kononova,M.	
TITLE		Cestrum yellow leaf curling virus promoters	
JOURNAL		Patent: WO 0173087-A 40 04 OCT-2001;	
FEATURES		Syngenta Participations AG (CH)	
source		Location/Qualifiers	
		1..1618	
		/organism="synthetic construct"	
		/db_xref="taxon:32630"	
		/note="artificial sequence Cmpc-synGPPI-nos expression cassette"	
BASE COUNT		468 a 383 c 400 g 367 t	







plants, removing transgenes from these plants or crops (e.g. food commodities), and restricting the distribution of transgenes within the environment

Claim 5: Fig 6; 60pp; English.

The invention describes an isolated excisable polynucleotide, which comprises a desired trait polynucleotide and a recombinase polynucleotide operably linked to a promoter, all flanked by a pair of directly oriented recombination sites where the recombinase activity is regulatable. The self-excising polynucleotide is useful for producing transgenic plants, particularly for removing all transgenic nucleic acid sequences that encode a gene product from the transgenic plant, thus restoring the original genetic configuration of the genome. The polynucleotide is also useful in methods for eliminating unwanted nucleic acids in agricultural food products and for preventing the escape of certain transgenic traits into the environment. This sequence encodes the novel *phic31* integrase *phic31intn*, created for use in the self-excising polynucleotide described in the invention.

Sequence 2031 BP: 451 A; 562 C; 640 G; 378 T; 0 other;

Query Match 100.0%; Score 2031; DB 24; Length 2031;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 2031; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCACAAGGGGTTGACCGGGGTGATACGTAAGTTCTGCTTACCTTGTATATA 60  
 DB 1 ATGGCACAAGGGGTTGACCGGGGTGATACGTAAGTTCTGCTTACCTTGTATATA 60  
 QY 61 TATATAATTAATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 120  
 DB 61 TATATAATTAATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 120  
 QY 121 AAAGATGTGATATAGCAATTCCTTCTGATTTATTAAGTGTGATATTTAATTTAATTT 180  
 DB 121 AAAGATGTGATATAGCAATTCCTTCTGATTTATTAAGTGTGATATTTAATTTAATTT 180  
 QY 181 ATACTTTTATATATATAGCAAAATTTCTTATGATGACAGTACCGGGTCTTATGACAC 240  
 DB 181 ATACTTTTATATATATAGCAAAATTTCTTATGATGACAGTACCGGGTCTTATGACAC 240  
 QY 241 CGTCACTGCGGCGAGCGAGATTCGAGCGAGCAAGCCAGCAGCAGCGTACGCGC 300  
 DB 241 CGTCACTGCGGCGAGCGAGATTCGAGCGAGCAAGCCAGCAGCAGCGTACGCGC 300  
 QY 301 AAGCAAGACAAGGGCGGCGACCTTCAGCGGAGTGAAGCGGAGCGGCGGTTACAG 360  
 DB 301 AAGCAAGACAAGGGCGGCGACCTTCAGCGGAGTGAAGCGGAGCGGCGGTTACAG 360  
 QY 361 TTGCTGGGCAATTCAGCAAGCGCGGCGAGTTCGCGGCGGCGGAGCGGCGCGC 420  
 DB 361 TTGCTGGGCAATTCAGCAAGCGCGGCGAGTTCGCGGCGGCGGAGCGGCGCGC 420  
 QY 421 GAGTTGAGACGATCTCTGACAGATGCGCGCGGCGGCGCTCAACATGATCTATTTGAT 480  
 DB 421 GAGTTGAGACGATCTCTGACAGATGCGCGCGGCGGCGCTCAACATGATCTATTTGAT 480  
 QY 481 GAGTTGAGACGATCTCTGAGAGCGCGGCGAGTTCGCGGCGGCGGAGCGGCGG 540  
 DB 481 GAGTTGAGACGATCTCTGAGAGCGCGGCGAGTTCGCGGCGGCGGAGCGGCGG 540  
 QY 541 CTCGCGCTGGGCGGTGACGATTTCTTCACTGAGAGCGGCGGCTTCCGCGAGGAACGTC 600  
 DB 541 CTCGCGCTGGGCGGTGACGATTTCTTCACTGAGAGCGGCGGCTTCCGCGAGGAACGTC 600  
 QY 601 ATGAGACGATCTCTGATATATGCGGCTCGAGCGGCTGCAACAAGAAATCTTCCGTAG 660  
 DB 601 ATGAGACGATCTCTGATATATGCGGCTCGAGCGGCTGCAACAAGAAATCTTCCGTAG 660  
 QY 661 TCGGGAAGATCTCTGACAGCAAGAACTTCAGCGCGAATTTGGGGGTACTGCGGCGG 720  
 DB 661 TCGGGAAGATCTCTGACAGCAAGAACTTCAGCGCGAATTTGGGGGTACTGCGGCGG 720

QY 721 AAGCGCCTTACGCGCTTCGACGCTTTGTCGAGACGAGAGAGATCAACGCCAAGCGCGA 780  
 DB 721 AAGCGCCTTACGCGCTTCGACGCTTTGTCGAGACGAGAGAGATCAACGCCAAGCGCGA 780  
 QY 781 ATGTCATATGCTGATCAACAAAGCTTGGGCACTGCAACCTCCCTTACCGGACCTTC 840  
 DB 781 ATGTCATATGCTGATCAACAAAGCTTGGGCACTGCAACCTCCCTTACCGGACCTTC 840  
 QY 841 GAGTTGAGAGCGGCGATATCCGTTGGTGGCTGAGATCAAGACGACAAACACTT 900  
 DB 841 GAGTTGAGAGCGGCGATATCCGTTGGTGGCTGAGATCAAGACGACAAACACTT 900  
 QY 901 CCCTTAAGCGGGGCGATCAAGCGGCGATTCACCGGCGACATCAAGGGCTTTGTAAG 960  
 DB 901 CCCTTAAGCGGGGCGATCAAGCGGCGATTCACCGGCGACATCAAGGGCTTTGTAAG 960  
 QY 961 CGCATGAGCGCTGACGCGCTGCGACCGGGGCGAGACGATTTGGGAAGAACCGCTTCA 1020  
 DB 961 CGCATGAGCGCTGACGCGCTGCGACCGGGGCGAGACGATTTGGGAAGAACCGCTTCA 1020  
 QY 1021 AAGCGCTGGGAGCGGCGCAACCGTTATGCGAATCTTTCGGGACCGGATTTCCGGGCTTC 1080  
 DB 1021 AAGCGCTGGGAGCGGCGCAACCGTTATGCGAATCTTTCGGGACCGGATTTCCGGGCTTC 1080  
 QY 1081 GCGGCTGAGGTGATCTACAAAGAAAGAGCGGAGCGGCGGCGGCGGCGGAT 1140  
 DB 1081 GCGGCTGAGGTGATCTACAAAGAAAGAGCGGAGCGGCGGCGGCGGCGGAT 1140  
 QY 1141 TACCGATTCAGCGCGGCGGCGATCACGCTCCGCGGCTGAGCTTATTCGCGGCGATC 1200  
 DB 1141 TACCGATTCAGCGCGGCGGCGATCACGCTCCGCGGCTGAGCTTATTCGCGGCGATC 1200  
 QY 1201 ATCGAGCGCGGCGGAGGATGATGAGCTTCAAGGCTGTTGAGCGGAGCGGCGGCGAAG 1260  
 DB 1201 ATCGAGCGCGGCGGAGGATGATGAGCTTCAAGGCTGTTGAGCGGAGCGGCGGCGAAG 1260  
 QY 1261 GGGCTTTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1320  
 DB 1261 GGGCTTTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1320  
 QY 1321 GCGGCTGAGCTTCTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1380  
 DB 1321 GCGGCTGAGCTTCTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1380  
 QY 1381 CGGAAGGTGCTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1440  
 DB 1381 CGGAAGGTGCTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1440  
 QY 1441 GCGGCTGAGCTTCTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1500  
 DB 1441 GCGGCTGAGCTTCTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1500  
 QY 1501 GACGAGAGAGCTTGGGCGGCTTCTGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1560  
 DB 1501 GACGAGAGAGCTTGGGCGGCTTCTGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1560  
 QY 1561 GCGGCTGAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1620  
 DB 1561 GCGGCTGAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1620  
 QY 1621 GCGGCTGAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1680  
 DB 1621 GCGGCTGAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1680  
 QY 1681 AAGCACTTCGGAAGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1740  
 DB 1681 AAGCACTTCGGAAGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1740  
 QY 1741 CTTCGCAACTTGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1800  
 DB 1741 CTTCGCAACTTGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1800

ID	Sequence	Name
Oy	1801 GAGCCGACGGCTGACGCCGACCAGCGCCCTTAATCTGTGTGGGGGGCCCCGCATAGACGAC	Phl C31 integrase; recombinase; transgenic; plant;
Dd	1801 GACCGCGACGGTGAACCCGACCGCGCCCTAACTCGTGTTGGGGGGCGCGTCAGTAGCGAC	agricultural food product; self-existing polynucleotide;
Oy	1861 AAGCGCGTTCGTCGGGGCTCTTCGTGAGCAAGATCGTTGTCAAGATCGACTACGGGC	plasmid; ds; pBBS EM051; T-DNA; cyclic; circular.
Dd	1861 AAGCGCGTTCGTCGGGGCTCTTCGTGAGCAAGATCGTTGTCAAGATCGACTACGGGC	Synthetic.
Oy	1921 AGGGGGCAGGGAMCGCCCATCGAAGAACCGCGCTTCGATCACTGAGGGCGAAGCGCGAC	Key
Dd	1921 .AGGGGGCAGGGAMCGCCCATCGAAGAACCGCGCTTCGATCACTGAGGGCGAAGCGCGAC	repeat_unit
Oy	1981 GACGACGACGAAGACGACGCCCCAGAGCGGACGACGAGGAGAGAGAGAGAGAGAGAG	misc_feature
Dd	1981 GACGACGACGAAGACGACGCCCCAGAGCGGACGACGAGGAGAGAGAGAGAGAGAGAG	terminator
Oy	1991 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	CDS
Dd	1991 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	promoter
Oy	2031 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	promoter
Dd	2031 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Oy	2091 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Dd	2091 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Oy	2151 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Dd	2151 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Oy	2211 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Dd	2211 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Oy	2271 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Dd	2271 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Oy	2331 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Dd	2331 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Oy	2391 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Dd	2391 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Oy	2451 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Dd	2451 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Oy	2511 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Dd	2511 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Oy	2571 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Dd	2571 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Oy	2631 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Dd	2631 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Oy	2691 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Dd	2691 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Oy	2751 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Dd	2751 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Oy	2811 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Dd	2811 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Oy	2871 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Dd	2871 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Oy	2931 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Dd	2931 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Oy	2991 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Dd	2991 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Oy	3051 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Dd	3051 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Oy	3111 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Dd	3111 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Oy	3171 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Dd	3171 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Oy	3231 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Dd	3231 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Oy	3291 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Dd	3291 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Oy	3351 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Dd	3351 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Oy	3411 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Dd	3411 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Oy	3471 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Dd	3471 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator
Oy	3531 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	terminator

```

FT FT /label= Super_Promoter
FT CDS 8413..9405
FT /tag= m
FT /product= "ergFP::INT"
FT terminator 9411..9677
FT /*tag= n
FT /label= Nopaline-synthase_terminator
FT misc_feature 9690..9728
FT /*tag= o
FT /label= atp
FT /note= "Recombinase target site"
FT repeat_unit 9735..9880
FT /*tag= p
FT /function=
FT "The right T-DNA border and the left T-DNA
FT border represent the end of the
FT self-excising polynucleotide"
PN MO200216609-A2.
XX 28-FEB-2002.
XX PD
XX PE
XX PF
XX PG
XX PH
XX PI
XX PJ
XX PK
XX PL
XX PM
XX PN
XX PO
XX PP
XX PQ
XX PR
XX PS
XX PT
XX PU
XX PV
XX PW
XX PX
XX PY
XX PZ
XX QA
XX QB
XX QC
XX QD
XX QE
XX QF
XX QG
XX QH
XX QI
XX QJ
XX QK
XX QL
XX QM
XX QN
XX QO
XX QQ
XX QR
XX QS
XX QT
XX QU
XX QV
XX QW
XX QX
XX QY
XX QZ
XX RA
XX RB
XX RC
XX RD
XX RE
XX RF
XX RG
XX RH
XX RI
XX RJ
XX RK
XX RL
XX RM
XX RN
XX RO
XX RP
XX RQ
XX RR
XX RS
XX RT
XX RU
XX RV
XX RW
XX RX
XX RY
XX RZ
XX SA
XX SB
XX SC
XX SD
XX SE
XX SF
XX SG
XX SH
XX SI
XX SJ
XX SK
XX SL
XX SM
XX SN
XX SO
XX SP
XX SQ
XX SR
XX SS
XX ST
XX SU
XX SV
XX SW
XX SX
XX SY
XX SZ
XX TA
XX TB
XX TC
XX TD
XX TE
XX TF
XX TG
XX TH
XX TI
XX TJ
XX TK
XX TL
XX TM
XX TN
XX TO
XX TP
XX TP
XX TR
XX TS
XX TT
XX TU
XX TV
XX TV
XX TW
XX TX
XX TY
XX TZ
XX UA
XX UB
XX UC
XX UD
XX UE
XX UF
XX UG
XX UH
XX UI
XX UJ
XX UK
XX UL
XX UM
XX UN
XX UN
XX UO
XX UP
XX UQ
XX UR
XX US
XX UT
XX UV
XX UV
XX UW
XX UX
XX UY
XX UZ
XX VA
XX VB
XX VC
XX VD
XX VE
XX VF
XX VG
XX VH
XX VI
XX VJ
XX VK
XX VL
XX VM
XX VN
XX VO
XX VP
XX VQ
XX VR
XX VS
XX VT
XX VU
XX VV
XX VW
XX VX
XX VY
XX VZ
XX WA
XX WB
XX WC
XX WD
XX WE
XX WF
XX WG
XX WH
XX WI
XX WJ
XX WK
XX WL
XX WM
XX WN
XX WO
XX WP
XX WQ
XX WR
XX WS
XX WT
XX WU
XX WV
XX WW
XX WX
XX WY
XX WZ
XX XA
XX XB
XX XC
XX XD
XX XE
XX XF
XX XG
XX XH
XX XI
XX XJ
XX XK
XX XL
XX XM
XX XN
XX XO
XX XP
XX XQ
XX XR
XX XS
XX XT
XX XU
XX XV
XX XV
XX XW
XX XX
XX XY
XX YZ
XX ZA
XX ZB
XX ZC
XX ZD
XX ZE
XX ZF
XX ZG
XX ZH
XX ZI
XX ZJ
XX ZK
XX ZL
XX ZM
XX ZN
XX ZO
XX ZP
XX ZQ
XX ZR
XX ZS
XX ZT
XX ZU
XX ZV
XX ZV
XX ZW
XX ZX
XX ZY
XX ZZ

```

Qy	1321	GCCTCATGACTGTGAACGCGGGGGAGATTCGATCAAGACTCTTACCGCTTGCCCTGC	1380
Db	6297	GCCGTATGACTTGTGAAACGCGGGGGAGATTCATCAAGACTCTTACCGCTTGCCCTGC	6356
Qy	1381	CGAAGTGGTGTGACCCCTCCGACCTGGGGACACAGGAAGGCACTGCMACTCAGCATG	1440
Db	6357	CGAAGTGGTGTGACCCCTCCGACCTGGGGACACAGGAAGGCACTGCMACTCAGCATG	6416
Qy	1441	GGGGCACTGCAACAAATTCGTTGGCGGAACGCATCTTCAACAAGATCAGGCACGCCAAGGC	1500
Db	6417	GGGGCACTGCAACAAATTCGTTGGCGGAACGCATCTTCAACAAGATCAGGCACGCCAAGGC	6476
Qy	1501	GACGAAGAGAGCTTGGCGCTTCTGTGGGAAGCGCCGACGCTTGCGCAAGTCACTAG	1560
Db	6477	GACGAAGAGAGCTTGGCGCTTCTGTGGGAAGCGCCGACGCTTGCGCAAGTCACTAG	6536
Qy	1561	GGCGCTTGGAAGACGCGCGGAACGCGGCAACCTTGTGGGAGCGCGCCGACGCCCTGAAC	1620
Db	6537	GGCGCTTGGAAGAGCGCGCGGAACGCGGCAACCTTGTGGGAGCGCGCCGACGCCCTGAAC	6596
Qy	1621	GCCCTTGAAGAGCTGTACGAAGAAGCGCGCGGACGCGCTTACGAGGACCGCTTGGGAGG	1680
Db	6597	GCCCTTGAAGAGCTGTACGAAGAAGCGCGCGGACGCGCTTACGAGGACCGCTTGGGAGG	6656
Qy	1681	AAGCACTTCCGGAAGCAACAGGCAAGCGCTGACGCTCCGGCAGCAAGGGGCGGAAGAGCGG	1740
Db	6657	AAGCACTTCCGGAAGCAACAGGCAAGCGCTGACGCTCCGGCAGCAAGGGGCGGAAGAGCGG	6716
Qy	1741	CTTGCGCAACTTGAAGCGCGCGCGCAAGCGCTTCCCTTGACCAATGGTTCCCGGA	1800
Db	6717	CTTGCGCAACTTGAAGCGCGCGCGCAAGCGCTTCCCTTGACCAATGGTTCCCGGA	6776
Qy	1801	GACGCCAAGCTGACCGACCGCGCCCTTAAGTGTGTGGGGGCGCGCTCAGTAGACAGAC	1860
Db	6777	GACGCCAAGCTGACCGACCGCGCCCTTAAGTGTGTGGGGGCGCGCTCAGTAGACAGAC	6836
Qy	1861	AAGCGCTGTTCGTCGGGGCTCTTCGTAGACAGATCGTTGTCAGAAAGTCAGTACGAGGC	1920
Db	6837	AAGCGCTGTTCGTCGGGGCTCTTCGTAGACAGATCGTTGTCAGAAAGTCAGTACGAGGC	6896
Qy	1921	AGGGGCGAAGGAAGCCCATCGAAGACGCGCTTCGATCAGCTGAGGCGGAAGCCGCGAC	1980
Db	6897	AGGGGCGAAGGAAGCCCATCGAAGACGCGCTTCGATCAGCTGAGGCGGAAGCCGCGAC	6956
Qy	1981	GACACACAGCAAGACGAGCGCCAGGACGCGCAAGAAAGCTAAGGGGGGTAG	2031
Db	6957	GACACACAGCAAGACGAGCGCCAGGACGCGCAAGAAAGCTAAGGGGGGTAG	7007
RESULT 3			
ABK12556			
1D	ABK12556	standard; DNA; 2031 BP.	
XX	ABK12556;		
XX	18-JUN-2002	(first entry)	
XX	DNA	encoding novel phi C31 integrase, phiC31Int*INT.	
XX	Phi C31 integrase; recombinase; transgenic; plant;		
XX	agricultural food product; self-excising polynucleotide;		
XX	gene; ds. phiC31Int*INT.		
XX	Synthetic.		
XX	WO200216609-A2.		
XX	28-FEB-2002.		
XX	27-AUG-2001; 2001WO-US26738.		
XX	25-AUG-2000; 2000US-227961P.		



XX (BADI) BASF PLANT SCI GMBH.  
PA (MCKE/) MCKERSIE B.  
XX  
PI Mankin L:  
XX WPI: 2002-280939/32.  
DR  
XX New self-excising polynucleotides, useful for producing transgenic  
PT plants, removing transgenes from these plants or crops (e.g. food  
PT commodities), and restricting the distribution of transgenes within the  
PT environment.  
PS  
XX Claim 5: Fig 7: 60pp: English.  
XX  
CC The invention describes an isolated excisable polynucleotide, which  
CC comprises a desired trait polynucleotide and a recombinase polynucleotide  
CC operably linked to a promoter, all flanked by a pair of directly oriented  
CC recombination sites where the recombinase activity is regulatable. The  
CC self-excising polynucleotide is useful for producing transgenic plants,  
CC particularly for removing all transgenic nucleic acid sequences that  
CC encode a gene product from the transgenic plant, thus restoring the  
CC original genetic configuration of the genome. The polynucleotide is also  
CC useful in methods for eliminating unwanted nucleic acids in agricultural  
CC food products and for preventing the escape of certain transgenic traits  
CC into the environment. This sequence encodes the novel phiC31 integrase  
CC phiC31int-INT, created for use in the self-excising polynucleotide  
CC described in the invention.  
XX  
XX Sequence 2031 BP: 454 A: 558 C: 640 G: 379 T: 0 other:  
SQ  
Query Match 99.2% Score 2015: DB 24: Length 2031:  
Best Local Similarity 99.5%: Pred. No. 0: Mismatches 10: Indels 0: Gaps 0:  
Matches 2021: Conservative 0: Mismatches 10: Indels 0: Gaps 0:

QY 1 ATGCACAAGGGGTGTGACCGGGGTGATACGTAAAGTTTCTTCTACCTTTATATA 60  
DB 1 ATGCACAAGGGGTGTGACCGGGGTGATACGTAAAGTTTCTTCTACCTTTATATA 60  
QY 61 TATATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 120  
DB 61 TATATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 120  
QY 121 AAGAATAGTATATAGCAATTCCTTCTGATGTTTAAAGTGTATATTTAATTT 180  
DB 121 AAGAATAGTATATAGCAATTCCTTCTGATGTTTAAAGTGTATATTTAATTT 180  
QY 181 ATAACTTTCTAATATATGACCAAAATTTGTGATGTGACGAGTACCGGCTCTACGAC 240  
DB 181 ATAACTTTCTAATATATGACCAAAATTTGTGATGTGACGAGTACCGGCTCTACGAC 240  
QY 241 CGTCAGTGGCGGACGCGGAGAAATTCGACGCGAGCAAGCCGACACAGCTGACGCC 300  
DB 241 CGTCAGTGGCGGACGCGGAGAAATTCGACGCGAGCAAGCCGACACAGCTGACGCC 300  
QY 301 AACGAAGACAAGGCGGCGGACCTTACGCGCGAAGTCGAGCGGAGCGGGCGGTTCAAG 360  
DB 301 AACGAAGACAAGGCGGCGGACCTTACGCGCGAAGTCGAGCGGAGCGGGCGGTTCAAG 360  
QY 361 TTTCGTGGGCAATTCAGGGAAGCGCGGACGCTGCGGTTTCGAGACGCGGAGCGCCG 420  
DB 361 TTTCGTGGGCAATTCAGGGAAGCGCGGACGCTGCGGTTTCGAGACGCGGAGCGCCG 420  
QY 421 GAGTTTGAAGCGATCTGAGCAATTCGCGCGCGGCGGCTTACATGATCTTGTCTAT 480  
DB 421 GAGTTTGAAGCGATCTGAGCAATTCGCGCGCGGCGGCTTACATGATCTTGTCTAT 480  
QY 481 GACGTGTGCGGCTTCTCGCGCTGAAGGTCATGACGCGGATTCGATGCTGGAATTTG 540  
DB 481 GACGTGTGCGGCTTCTCGCGCTGAAGGTCATGACGCGGATTCGATGCTGGAATTTG 540  
QY 541 CTCGCGCTGGGCGTACGATTTGTTTCACTCAGGAAGCGCTTCTCGGAGGGAACGTC 600  
DB 541 CTCGCGCTGGGCGTACGATTTGTTTCACTCAGGAAGCGCTTCTCGGAGGGAACGTC 600

DB 541 CTCGCGCTGGGCGTACGATTTGTTTCACTCAGGAAGCGCTTCTCGGAGGGAACGTC 600  
QY 601 ATGGAACCTGATACCTGATTAATGCGGCTGCGACCGGTCCGCAAGAATCTTCGCGAG 660  
DB 601 ATGGAACCTGATACCTGATTAATGCGGCTGCGACCGGTCCGCAAGAATCTTCGCGAG 660  
QY 661 TCGGCGAAGATTCGAGACGCAAGAACCTTACGCGCGAATTTGGGCGGCTGCGCGG 720  
DB 661 TCGGCGAAGATTCGAGACGCAAGAACCTTACGCGCGAATTTGGGCGGCTGCGCGG 720  
QY 721 AAGGCGCTTACGCGCTTGTGAGAGACGAGATTCACGCGCAAGCGCGCA 780  
DB 721 AAGGCGCTTACGCGCTTGTGAGAGACGAGATTCACGCGCAAGCGCGCA 780  
QY 781 ATGTCGAATGCTGATCAACAAGCTTGGGCACTGACACATCCCTTACCGGACCTTC 840  
DB 781 ATGTCGAATGCTGATCAACAAGCTTGGGCACTGACACATCCCTTACCGGACCTTC 840  
QY 841 GAGTTTGAAGCGGAGTATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
DB 841 GAGTTTGAAGCGGAGTATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
QY 901 CCCCTTCAAGCGGCGGAGTCAAGCCGCTTACACCGGCGAGATCACGGGCTTGTAG 960  
DB 901 CCCCTTCAAGCGGCGGAGTCAAGCCGCTTACACCGGCGAGATCACGGGCTTGTAG 960  
QY 961 CGCATGAGCGTGAAGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020  
DB 961 CGCATGAGCGTGAAGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020  
QY 1021 AGCGGCTGGGACCGGCGCAACCGCTTATGCAATCTTCCGCGGCGGCTTATGCGGCTTC 1080  
DB 1021 AGCGGCTGGGACCGGCGCAACCGCTTATGCAATCTTCCGCGGCGGCTTATGCGGCTTC 1080  
QY 1081 GCCGCTGAGTATCTACAAGAAAGCCGAGCGGCGGCGGCGGCGGCGGCGGCGGCT 1140  
DB 1081 GCCGCTGAGTATCTACAAGAAAGCCGAGCGGCGGCGGCGGCGGCGGCGGCGGCT 1140  
QY 1141 TACCGGCTTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200  
DB 1141 TACCGGCTTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200  
QY 1201 ATGGAACCGGCTGAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 1260  
DB 1201 ATGGAACCGGCTGAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 1260  
QY 1261 GGGCTTTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1320  
DB 1261 GGGCTTTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1320  
QY 1321 GCGGCTATGATCTTCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1380  
DB 1321 GCGGCTATGATCTTCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1380  
QY 1381 CGGAAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1440  
DB 1381 CGGAAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1440  
QY 1441 GCGGCACTGACAGTCTTCTGCGGAAGCATCTTCAACAAGATCAGGACGCGCGAAGGC 1500  
DB 1441 GCGGCACTGACAGTCTTCTGCGGAAGCATCTTCAACAAGATCAGGACGCGCGAAGGC 1500  
QY 1501 GACGAAGAGAGCTTGGGCTTCTGAGGAAGCGGCGGCGGCGGCGGCGGCGGCGG 1560  
DB 1501 GACGAAGAGAGCTTGGGCTTCTGAGGAAGCGGCGGCGGCGGCGGCGGCGGCGG 1560  
QY 1561 GCGGCTGAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1620  
DB 1561 GCGGCTGAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1620  
QY 1621 GCCCTTGAAGAGCTTGAAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1680  
DB 1621 GCCCTTGAAGAGCTTGAAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1680



QY	813	CTGACACACTCCCTTACCGGACCCTTGCAGTTTCAGACCCGACGTAATCCGCTGCTG	872
Db	1601	CTTGACCACTCCCTTACCGGACCCTTGCAGTTTCAGACCCGACGTAATCCGCTGCTG	1660
QY	873	GCAGTAGTCAAGACGGCAAAACACGTCCTTCAAGCCGGGCACTCAAAAGCCGATTC	932
Db	1661	GCCTGAGTCAAGACGGCAAAACACCTTCCCTTCAAGCCGGGCACTCAAAAGCCGATTC	1720
QY	933	CCCGGCGAGCTCACGGGCGCTTGTAAAGCGATGACGCTGACGCGCTGACCGACCCGGG	992
Db	1721	CCCGGCGAGCATACAGGGGCTTGTAAAGCGATGACGCGCTGACGCGCTGACCGACCCGGG	1780
QY	993	CGAAGCATTTGGGAAGAAAGCCGCTTCAAGCGCTGGAACCCGGCAACGCTTATGCGAT	1052
Db	1781	CGAGCATTTGGGAAGAAAGCCGCTTCAAGCGCTGGAACCCGGCAACGCTTATGCGAT	1840
QY	1053	CGTTCCGGGACCCGGGCTTTCGGGGCTTTCGGCGCTGAGGATCTCAAGAAAGAGCGGA	1112
Db	1841	CGTTCCGGGACCCGGGCTTTCGGGGCTTTCGGCGCTGAGGATCTCAAGAAAGAGCGGA	1900
QY	1113	CGGACGCGCGACCAAGAAAGTTAGAGGTTACCGCATTCAGCCGCAACCCGATACGCTCG	1172
Db	1901	CGGACGCGCGACCAAGAAAGTTAGAGGTTACCGCATTCAGCGCGCAACCCGATACGCTCG	1960
QY	1173	GCCGCTGAGACTTATTTGCGGACCATCATCGACCCGCTGAGTGATATGACTTCAGC	1232
Db	1961	GCCGCTGAGACTTATTTGCGGACCATCATCGACCCGCTGAGTGATATGACTTCAGC	2020
QY	1233	GTCGTTGAGAGCGAGGGGGGGGCGCAAGGGGCTTCCGGGGGCAAGCATTTCTGTCGC	1292
Db	2021	GTCGTTGAGAGCGAGGGGGGGGCGCAAGGGGCTTCCGGGGGCAAGCATTTCTGTCGC	2080
QY	1293	CATGAGACAAGCTGACTGCGCATGTTGGGCGCGCTCATGATTCGAAGCCGCGGAAGATC	1352
Db	2081	CATGAGACAAGCTGACTGCGCATGTTGGGCGCGCTCATGATTCGAAGCCGCGGAAGATC	2140
QY	1333	GATCAAGACTCTTACCGCTGCGCGTGGCGGGAAGTGTCGACCCGCTCGCACCTGGGCA	1412
Db	2141	GATCAAGACTCTTACCGCTGCGCGTGGCGGGAAGTGTCGACCCGCTCGCACCTGGGCA	2200
QY	1413	GCAGGAAGGCACTGCAAGCTTCAGAGTGGCGGCACTGCAAGTTCGTTGGGGAAGCAT	1472
Db	2201	GCAGGAAGGCACTGCAAGCTTCAGAGTGGCGGCACTGCAAGTTCGTTGGGGAAGCAT	2260
QY	1473	CTTCAACAAGATGACGACCCGCAAGGCGCAAGAGAGACTTGGCGCTTCTGTGGGAAGC	1532
Db	2261	CTTCAACAAGATGACGACCCGCAAGGCGCAAGAGAGACTTGGCGCTTCTGTGGGAAGC	2320
QY	1533	CGCCGACGCTTCGGCAAGCTCACTAGGCGGCTGGAAGAGCGGCGCAAGCGGCGAACC	1592
Db	2321	CGCCGACGCTTCGGCAAGCTCACTAGGCGGCTGGAAGAGCGGCGCAAGCGGCGAACC	2380
QY	1593	TGTTTGGGAGGGCCCGCAACCCCTGAAAGCGCTTGAAGAGCTGTACGAAGACCGCGGCG	1652
Db	2381	TGTTTGGGAGGGCCCGCAACCCCTGAAAGCGCTTGAAGAGCTGTACGAAGACCGCGGCG	2440
QY	1653	AGGCGCGTACGAGCGACCCGCTTGGGCAAGAGCACTTCGGAAGCAAGGCAAGCGGCTGAC	1712
Db	2441	AGGCGCGTACGAGCGACCCGCTTGGGCAAGAGCACTTCGGAAGCAAGGCAAGCGGCTGAC	2500
QY	1713	GCTCCGCGAGCAAGGGCGGAGAGGCGCTTGCAGAACTTGAAGCCGCGCAAGCCCGCA	1772
Db	2501	GCTCCGCGAGCAAGGGCGGAGAGGCGCTTGCAGAACTTGAAGCCGCGCAAGCCCGCA	2560
QY	1773	GCTTTCCTTGTACCATATGTTCCCGCAAGACGCGCTGACCCGACCGGCGCTTAAGTC	1832
Db	2561	GCTTTCCTTGTACCATATGTTCCCGCAAGACGCGCTGACCCGACCGGCGCTTAAGTC	2620
QY	1833	GTCGTCGGGGGGCGCGCTACATAGACGCAAGCGCGTTCGTCGGGGCTTCTGTAGCAAA	1892
Db	2621	GTCGTCGGGGGGCGCGCTACATAGACGCAAGCGCGTTCGTCGGGGCTTCTGTAGCAAA	2680
QY	1893	GATCGTTGTCAAGATGCACTACGGGCAAGGGGCGAGGGCAAGCCCATCTGAAGAGCGGC	1952

Db	2681	GATCGCTTGTCAACGAAGTGACTGACTACGGGCAGGGGGCAGGAACGCCCATCGAAGAAGCCGCG	2740
Oy	1953	TTCGATTCAACGTGGGGGAAAGCCGCCGACCGACGACGACGACGAAGACGACCAGGACGAC	2012
Db	2741	TTTCGATTCAACGTGGGGGAAAGCCGCCGACCGACGACGACGACGAAGACGACCAGGACGAC	2800
-Oy	2013	GGAGACGCTAGCGCGCGTAG	2031
Db	2801	GGAGACGCTAGCGCGCGTAG	2819
RESULT 5			
AHH74879			
ID	AHH74879	standard; DNA; 1842 BP.	
XX	AHH74879;		
AC			
XX			
DT	29-OCT-2001	(first entry)	
XX			
DE			
XX			
KM	Nucleotide sequence of a wildtype bacteriophage thetaC31 integrase.		
KW	Recombinase; genetic engineering; monogenic disorder; ADA deficiency;		
KW	cystic fibrosis; familial-hypercholesterolaemia; haemophilia;		
KW	chronic granulomatous disease; Duchenne's muscular dystrophy;		
KW	Fanconi's disease; anemia; sickle-cell anemia; Gaucher's disease;		
KW	Hunter's syndrome; X-linked severe combined immunodeficiency; SCID;		
KW	infectious disease; acquired disorder; tumour; cancer; integrase; ss.		
OS	Bacteriophage thetaC31.		
PN	WO200161049-A1.		
XX			
PD	23-AUG-2001.		
XX			
Pf	16-FEB-2001; 2001WO-US05269.		
PR	18-FEB-2000; 2000US-0183759.		
PA	(STRD ) UNIV DELAND STANFORD JUNIOR.		
PI	Calos MP, Sclimentl CR;		
DR	WPI: 2001-522610/57.		
PT			
PT	Identifying altered recombinase, involves hybridizing recombination		
PT	sites in cells having specific coding sequence, transformed with		
PT	altered recombinase gene, and isolating cells having product of the		
PT	sequence -		
PS	Disclosure: Fig 4; 101pp; English.		
XX			
XX	The specification describes a method for identifying altered		
CC	recombinases. The method comprises transforming cells with a first		
CC	plasmid comprising two recombination sites and a coding sequence of		
CC	interest, and a second plasmid encoding an altered recombinase; allowing		
CC	recombination of the two recombination sites using the altered		
CC	recombinase; screening and isolating transformed cells comprising the		
CC	product of the sequence of interest; and identifying the altered		
CC	recombinase. The altered recombinase is useful for site-specifically		
CC	integrating a polynucleotide sequence of interest in a genome of a cell.		
CC	It is also useful in genetic engineering of chromosomes of higher cells,		
CC	and for the generation of transgenic cells, tissues, plants and animals.		
CC	The altered recombinase is useful for treating monogenic disorders,		
CC	e.g. ADA deficiency, cystic fibrosis, familial-hypercholesterolaemia,		
CC	hemophilia, chronic granulomatous disease, Duchenne's muscular dystrophy,		
CC	haemophilia, Fanconi's disease, sickle-cell anemia, Gaucher's disease,		
CC	Hunter's syndrome and X-linked severe combined immunodeficiency (SCID),		
CC	infectious diseases including viral and bacterial infections, acquired		
CC	disorders including solid tumours and hematopoietic cancers such as		
CC	leukaemias and lymphomas, and other cancers. The present sequence		
CC	encodes a wild type bacteriophage thetaC31 integrase.		
XX			

Sequence 1842 BP; 387 A; 547 C; 616 G; 292 T; 0 other;

Query Match 89.2%; Score 1811; DB 22; Length 1842;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1814; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 213 GATGTGACGAGTACGGGGTGTCTTACGACCTGACGCGGAGCGGAGAAATTCGAGCCG 272  
DB 24 GGTGGACACGATACGGGGTGTCTTACGACCTGACGCGGAGCGGAGAAATTCGAGCCG 83  
QY 273 AGCAAGCCAGACGACAGCGGTACCGCAAGCAAGAGCGGCGGACCTTCAGCGCGA 332  
DB 84 AGCAAGCCAGACGACAGCGGTACCGCAAGCAAGAGCGGCGGACCTTCAGCGCGA 143  
QY 333 AGTCGACGCGGACGCGGCGGTTCAGGTTGTCGCGGATTCAGCGAAGCGCGGCGAC 392  
DB 144 AGTCGACGCGGACGCGGCGGTTCAGGTTGTCGCGGATTCAGCGAAGCGCGGCGAC 203  
QY 393 GTGCGGCTTGGGACGCGGCGGAGCGCGGAGTTGGAAGCATCCTGGAAGATGCCGCG 452  
DB 204 GTGCGGCTTGGGACGCGGCGGAGCGCGGAGTTGGAAGCATCCTGGAAGATGCCGCG 263  
QY 453 CGGCGGCTCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 512  
DB 264 CGGCGGCTCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 323  
QY 513 GGAACGCGATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 572  
DB 324 GGAACGCGATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 383  
QY 573 GGAAGGCGTTCGCGGAGGGAAGCAATGACGATGATGATGATGATGATGATGATGATGAT 632  
DB 384 GGAAGGCGTTCGCGGAGGGAAGCAATGACGATGATGATGATGATGATGATGATGATGAT 443  
QY 633 CGCGTGCACAAAGAAATCTTCTGCTGAAGTCGCGGAGATTCGACAGCAAGAACTTGA 692  
DB 444 CGCGTGCACAAAGAAATCTTCTGCTGAAGTCGCGGAGATTCGACAGCAAGAACTTGA 503  
QY 693 GCGCGAATTTGGGCGGCTGAGTGTGCGGCGGGAAGCGCCTTACGCGCTTGAGCTTGTTCGA 752  
DB 504 GCGCGAATTTGGGCGGCTGAGTGTGCGGCGGGAAGCGCCTTACGCGCTTGAGCTTGTTCGA 563  
QY 753 GACGAGGAGATACACGCGACGCGCAATGTCATGTCGTCATCAACAAGCTTGGCGA 812  
DB 564 GACGAGGAGATACACGCGACGCGCAATGTCATGTCGTCATCAACAAGCTTGGCGA 623  
QY 813 CTCGACCATCTCCCTTACCGGACCTTTCGAGTTCGAGCCGCAAGTATCGGTGTGTG 872  
DB 624 CTCGACCATCTCCCTTACCGGACCTTTCGAGTTCGAGCCGCAAGTATCGGTGTGTG 683  
QY 873 GCGTGATCAAGACGACCAACACCTTCCCTTCAAGCGGCGGAGTCAGACCGCATTCGA 932  
DB 684 GCGTGATCAAGACGACCAACACCTTCCCTTCAAGCGGCGGAGTCAGACCGCATTCGA 743  
QY 933 CCGCGGCGACATACGCGGCGTTCGAGCGATGAGCGTTCGCGGCGGAGCGCGG 992  
DB 744 CCGCGGCGACATACGCGGCGTTCGAGCGATGAGCGTTCGCGGCGGAGCGCGG 803  
QY 993 CGAGACGATTTGGGAGGAGACCGGCTTCAAGCGCTTGGAGCCCGGCAACCGTTATCGAAT 1052  
DB 804 CGAGACGATTTGGGAGGAGACCGGCTTCAAGCGCTTGGAGCCCGGCAACCGTTATCGAAT 863  
QY 1053 CCTTGGGAGCGCGCTATTTGCGGCGTTCGCGCTGAGTGTATCTCAAGAGAAACCGGA 1112  
DB 864 CCTTGGGAGCGCGCTATTTGCGGCGTTCGCGCTGAGTGTATCTCAAGAGAAACCGGA 923  
QY 1113 CGGACGCGCGACCAAGAAATTTAGGCTTACCGCATTCAGCGCGACCGCATCAGCTCCG 1172  
DB 924 CGGACGCGCGACCAAGAAATTTAGGCTTACCGCATTCAGCGCGACCGCATCAGCTCCG 983  
QY 1173 GCGCGGTGAGCTTGTATGCGGACGATCATGAGCGCGCTAGTGTATGAGCTTCAGCG 1232  
DB 984 GCGCGGTGAGCTTGTATGCGGACGATCATGAGCGCGCTAGTGTATGAGCTTCAGCG 1043

QY 1233 GTGTTGGACGAGGCGGCGGCAAGGGGCTTTCCCGGGGCAAGCCATTGTGTCGG 1292  
DB 1044 GTGTTGGACGAGGCGGCGGCAAGGGGCTTTCCCGGGGCAAGCCATTGTGTCGG 1103  
QY 1293 CATGCAACAGCTTACTGTGAGTGTGGCGCGCATGACCTTCAAGCGGCGGGAACAATC 1352  
DB 1104 CATGCAACAGCTTACTGTGAGTGTGGCGCGCATGACCTTCAAGCGGCGGGAACAATC 1163  
QY 1353 GATCAAGAGCTTTCACCGCTGCCGTCGCGGAAGTGTGTGACCCGTCGCACTTGGGA 1412  
DB 1164 GATCAAGAGCTTTCACCGCTGCCGTCGCGGAAGTGTGTGACCCGTCGCACTTGGGA 1223  
QY 1413 GCACGAAGGACGTGCAAGGTGACATGCGGCGACTGCAAGTTCGTTGGCGAAGCAT 1472  
DB 1224 GCACGAAGGACGTGCAAGGTGACATGCGGCGACTGCAAGTTCGTTGGCGAAGCAT 1283  
QY 1473 CTTCAACAAGATCAGGACGCGCAAGGCGAGCAAGAGAGCTTGGCGCTTCTGTGGAGC 1532  
DB 1284 CTTCAACAAGATCAGGACGCGCAAGGCGAGCAAGAGAGCTTGGCGCTTCTGTGGAGC 1343  
QY 1533 CGCGGACGCTTGGCAAGCTCAGTGAAGCGCTGAGAAAGCGGCGGAGCAACCT 1592  
DB 1344 CGCGGACGCTTGGCAAGCTCAGTGAAGCGCTGAGAAAGCGGCGGAGCAACCT 1403  
QY 1593 TGTGGGAGCGCGCGACGCGCTGAAAGCGCTTGAAGAGCTGTAAGAACCGGCGGCG 1652  
DB 1404 TGTGGGAGCGCGCGACGCGCTGAAAGCGCTTGAAGAGCTGTAAGAACCGGCGGCG 1463  
QY 1653 AGCGCGTACGACGAGACCGCTTGGCAAGACCTTCGGAAGCAACAAGGCAAGCTGAC 1712  
DB 1464 AGCGCGTACGACGAGACCGCTTGGCAAGACCTTCGGAAGCAACAAGGCAAGCTGAC 1523  
QY 1713 GCTCCGGACGCAAGGGGCGGAGAGCGGCTTCCGCAACTTGAAGCGCGGCAAGCCGGA 1772  
DB 1524 GCTCCGGACGCAAGGGGCGGAGAGCGGCTTCCGCAACTTGAAGCGCGGCAAGCCGGA 1583  
QY 1773 GCTTCCCTTGAACAATGTTCCCGGAAGAGCGCAGCCTGACCGGCGGCTTAAATGTC 1832  
DB 1584 GCTTCCCTTGAACAATGTTCCCGGAAGAGCGCAGCCTGACCGGCGGCTTAAATGTC 1643  
QY 1833 GTGGTGGGCGCGCGCTGATGACGACAAAGCGCTTGTGTCGCGGCTTCTGTGACAA 1892  
DB 1644 GTGGTGGGCGCGCGCTGATGACGACAAAGCGCTTGTGTCGCGGCTTCTGTGACAA 1703  
QY 1893 GATCGTTGCAAGATGACGACGAGGCGGAGGCGGAGGAGGAGGAGGAGGAGGAGG 1952  
DB 1704 GATCGTTGCAAGATGACGACGAGGCGGAGGCGGAGGAGGAGGAGGAGGAGGAGG 1763  
QY 1953 TTGATACAGTGGGGAAGCGCGGCGGACCGGACGAGCAGCAGAGAGAGAGAGAGAG 2012  
DB 1764 TTGATACAGTGGGGAAGCGCGGCGGACCGGACCGGACGAGCAGCAGAGAGAGAGAG 1823  
QY 2013 GGAAGAGCTAGCGGCGTAG 2031.  
DB 1824 GGAAGAGCTAGCGGCGTAG 1842

# RESULT 6

AA006904

ID AA006904 standard; DNA; 3401 BP.

XX AC AA006904;

DT 06-MAR-1991 (first entry)

XX DE Sequence encoding site specific integrating function of actinomycete  
XX DE phage phi C31.  
XX KM Antibiotic production; Isovaleryl; spiramycin; ds.  
XX OS Page phi C31.





Db	924	CGGCACCCCGACACGAAAGTTAAGGGTTTACCCGATTCAGCCGACCCGATTCACGCTCCG	983
Qy	1173	GCCCGCTGACCTTGATTTGCGGACCGATCATCGAGCCCGCTGAAGTGTATGACTTCAGCG	1232
Dp	984	GCCCGGTGACACTTGATTTGCGGACCCGATCATTCGAGCCCGCTGAGTGTATGAGCTTCAGCG	1043
Qy	1233	GTGGTTGACGGGCAAGGGGGGGGGGCAAGGGGGCTTCCCGGGGGCAAGCCATTCTGTCCGC	1292
Dp	1044	GTGGTTGACGGGCAAGGGGGGGGGGCAAGGGGGCTTCCCGGGGGCAAGCCATTCTGTCCGC	1103
Qy	1293	CATGACAAAGCTGTACTGACGATGTGTGGCCCGCTCATGATCTTGTGAAGCCGCGGGAAAGATC	1352
Dp	1104	CATGACAAAGCTGTACTGACGATGTGTGGCCCGCTCATGATCTTGTGAAGCCGCGGGAAAGATC	1153
Qy	1353	GATTAAGGACTTTTACCCTGCTGCCGTGCGGGAAGGTGTGACACCCTCCGACCTGGGCA	1412
Dp	1164	GATTAAGGACTTTTACCCTGCTGCCGTGCGGGAAGGTGTGACACCCTCCGACCTGGGCA	1233
Qy	1413	GCACGAAGGCGCGGCAACCTCAGCAATGGGGGACCTCGACATCTTGTTGGGGAAAGCAT	1472
Dp	1224	GCACGAAGGCGCGGCAACCTCAGCAATGGGGGACCTCGACATCTTGTTGGGGAAAGCAT	1283
Qy	1473	CTTCAACAAGATCAGGACCGCCGGAAGCGCAGAAAGACGTTTGGCGCTTCTGTGGGAAGC	1532
Dp	1284	CTTCAACAAGATCAGGACCGCCGGAAGCGCAGAAAGACGTTTGGCGCTTCTGTGGGAAGC	1343
Qy	1533	CGCCCGGACGTTTGGGCAAGCTCACTGAGGCGCTTGAAGAAAGCGCGCAACGCGCAACT	1592
Dp	1344	CGCCCGGACGTTTGGGCAAGCTCACTGAGGCGCTTGAAGAAAGCGCGCAACGCGCAACT	1403
Qy	1593	TGTTGCGAGAGCGCGCAGCCCTTGAAGCGCCTTGAAGAGCTGTACGAAGAACCGCGCGCGC	1652
Dp	1404	TGTTGCGAGAGCGCGCAGCCCTTGAAGCGCCTTGAAGAGCTGTACGAAGAACCGCGCGCGC	1463
Qy	1653	AGGCGCGTACGACGAGACCCCTTGGCAGGAAGCATTCGGAAGCAAGCAGCGCTGAC	1712
Dp	1464	AGGCGCGTACGACGAGACCCCTTGGCAGGAAGCATTCGGAAGCAAGCAGCGCTGAC	1523
Qy	1713	GCTCGCGGACCAAGGGCGGGAAGAAGGGGCTTGGCGAATCTTGAAGCCCGCGAAGCCCGCA	1772
Dp	1544	GCTCGCGGACCAAGGGCGGGAAGAAGGGGCTTGGCGAATCTTGAAGCCCGCGCGAAGCCCGCA	1583
Qy	1773	GCTTCCCTTTGACCAATGTGTCCCGCAAGACGCGAGCTGACCCGACGCGGCGCTTAATTC	1832
Dp	1584	GCTTCCCTTTGACCAATGTGTCCCGCAAGACGCGAGCTGACCCGACGCGGCGCTTAATTC	1643
Qy	1833	GTGGTGGGGGGCGCGCTCAGTAGACGACAGAGCGGCTTGTGCGGGCTCTTCGTAGACAA	1892
Dp	1644	GTGGTGGGGGGCGCGCTCAGTAGACGACAGAGCGGCTTGTGCGGGCTCTTCGTAGACAA	1703
Qy	1893	GATGCTTGTACAGAAAGTCGACTACGGGGCAAGGGGGCAAGGGCAAGCCCATTCGAAAGCGGC	1952
Dp	1704	GATGCTTGTACAGAAAGTCGACTACGGGGCAAGGGGGCAAGGGCAAGCCCATTCGAAAGCGGC	1763
Qy	1953	TTTCATCAGCTGGGGCGAAAGCGCGGACCGACGACGACGACGAAAGACGACGCCGACGAGCGAC	2012
Dp	1764	TTTCATCAGCTGGGGCGAAAGCGCGGACCGGACCGACGACGACGAAAGACGACGCCGACGAGCGAC	1823
Qy	2013	GGAAGACGTAAGCGCGCTAG 2031	
Dp	1824	GCAAGACGTAAGCGCGCTAG 1842	

XX	AAH74881	standard; DNA; 1839 BP
XX	AAH74881;	
XX	29-Oct-2001	(first entry)
XX	Nucleotide sequence of altered	

XX Recombinase; genetic engineering; monogenic disorder; ADA deficiency;  
 KM cystic fibrosis; familial-hypercholesterolaemia; haemophilia;  
 KM chronic granulomatous disease; Duchenne's muscular dystrophy;  
 KM Fanconi's disease; anemia; sickle-cell anemia; Gaucher's disease;  
 KM Hunter's syndrome; X-linked severe combined immunodeficiency; SCID;  
 KM infectious disease; acquired disorder; tumour; cancer; ss.  
 XX Synthetic.  
 OS Bacteriophage.  
 XX  
 FH Key Location/Qualifiers  
 FT 1..1839  
 FT CDS /tag- a  
 FT /product= "altered recombinase 5C1"  
 FT /note= "no termination codon given."  
 XX  
 XX PN WO200161049-A1.  
 XX PD 23-AUG-2001.  
 XX PF 16-FEB-2001; 2001WO-US05269.  
 XX PR 18-FEB-2000; 2000US-0183759.  
 XX PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 XX PI Calos MP, Sclimenti CR;  
 XX DR WPI: 2001-522610/57.  
 XX DR P-PSDB: AMG63848.  
 XX  
 PT Identifying altered recombinase, involves hybridizing recombination  
 PT sites in cells having specific coding sequence, transformed with  
 PT altered recombinase gene, and isolating cells having product of the  
 PT sequence -  
 PS  
 XX  
 XX Disclosure: Fig 8: 101pp: English.  
 XX  
 CC The specification describes a method for identifying altered  
 CC recombinases. The method comprises transforming cells with a first  
 CC plasmid comprising two recombination sites and a coding sequence of  
 CC interest, and a second plasmid encoding an altered recombinase; allowing  
 CC recombination of the two recombination sites using the altered  
 CC recombinase; screening and isolating transformed cells comprising the  
 CC product of the sequence of interest; and identifying the altered  
 CC recombinase. The altered recombinase is useful for site-specifically  
 CC integrating a polynucleotide sequence of interest in a genome of a cell.  
 CC It is also useful in genetic engineering of chromosomes of higher cells,  
 CC and for the generation of transgenic cells, tissues, plants and animals.  
 CC The altered recombinase is useful for treating monogenic disorders,  
 CC e.g. ADA deficiency, cystic fibrosis, familial-hypercholesterolaemia,  
 CC anemia, chronic granulomatous disease, Duchenne's muscular dystrophy,  
 CC haemophilia, Fanconi's disease, sickle-cell anemia, Gaucher's disease,  
 CC Hunter's syndrome and X-linked severe combined immunodeficiency (SCID).  
 CC Infectious diseases including viral and bacterial infections, acquired  
 CC disorders including solid tumours and haematopoietic cancers such as  
 CC leukaemias and lymphomas, and other cancers. The present sequence  
 CC encodes an altered recombinase of the invention.  
 CC  
 XX  
 XX Sequence 1839 BP; 388 A; 547 C; 613 G; 291 T; 0 other;

0Y	213	GAGTGCAGGTAACGGGGTGGTTACACCGTCATCGCGGAGGCGCAGAAATTCGACGC	272
0Y	24	GTTGACACGTAAGGGGGTGGCTTAAGACCGTAGCGGCGAGCGCAAAATTCGACGC	83
0Y	273	AGCAAGCCACGACACAGCGTAGGCGCCAAAGACAGGCGGCGACCTTCAGCGCGA	332
Db	84	AGCAAGCCACGACACAGCGTAGGCGCCAAAGACAGGCGGCGACCTTCAGCGCGA	14



QY	1413	GCATCAAGGCAACGTTGCAACAGTACAGACATGAGGCGGCACTGCAAGATTGCTTGGCGAAGCGAT	14172
Db	1224	GCACGAAGGCAACGTTGCAACAGTACAGACATGAGGCGGCACTGCAAGATTGCTTGGCGAAGCGAT	1283
QY	1473	CTTCAACAAGTATCAAGGCAACGCGCCGAAGGCGACGAAAGAGACGTTTGGCGCTTCTTGGGAAGC	15323
Db	1284	CTTCAACAAGTATCAAGGCAACGCGCCGAAGGCGACGAAAGAGACGTTTGGCGCTTCTTGGGAAGC	13433
QY	1533	CGCCGAGCGTTTGGCGAAGGCTCACTGAGGCGCTTGAGAAAGAGCGGGGAAACGGCGAGACT	15922
Db	1344	CGCCGAGCGTTTGGCGAAGGCTCACTGAGGCGCTTGAGAAAGAGCGGGGAAACGGCGAGACT	14030
QY	1593	TGTTTGGAGAGCGCGCGCGAGCGCCCTGAACGCGCTTGAAGAGCTGTACGAAAGACCGCGCGGC	16522
Db	1404	TGTTTGGAGAGCGCGCGCGAGCGCCCTGAACGCGCTTGAAGAGCTGTACGAAAGACCGCGCGGC	14633
QY	1653	AGGCGCGTACAGAGGACCGCTTGGCGAAGGCACTTCGGAAGCAACAGCGACGCGCTGAC	17122
Db	1464	AGGCGCGTACAGAGGACCGCTTGGCGAAGGCACTTCGGAAGCAACAGCGACGCGCTGAC	15233
QY	1713	GCTCCGCGAGCAAGGCGGAGAGAGCGCTTGGCGAAGTGTGAAGCGCGCGAAGCGCGCGAA	17722
Db	1524	GCTCCGCGAGCAAGGCGGAGAGAGCGCTTGGCGAAGTGTGAAGCGCGCGAAGCGCGCGAA	15833
QY	1773	GCTTCCCTTGACCAATGTTCCCGCAAGCGCCGACGCTGACCCGACCGCGCCTTAAGTC	18322
Db	1584	GCTTCCCTTGACCAATGTTCCCGCAAGCGCCGACGCTGACCCGACCGCGCCTTAAGTC	16433
QY	1833	GTGCTGGGCGCGCGCGTCAAGTATGACACAAAGGCGCTTTCGCGGGCTTTCGTATGACAA	18922
Db	1644	GTGCTGGGCGCGCGCGTCAAGTATGACACAAAGGCGCTTTCGCGGGCTTTCGTATGACAA	17033
QY	1893	GATCGTGTGTACCAAGTCTACGAGGCGAGGCGGCAAGGCAAGCGCCCATCGAAGAGCGCGC	19522
Db	1704	GATCGTGTGTACCAAGTCTACGAGGCGAGGCGGCAAGGCAAGCGCCCATCGAAGAGCGCGC	17633
QY	1953	TTCGATACGTTGGGCGAAGGCGCGCGACGACGAGACGAAAGCAAGCGCGCGACGAGCGGAC	20122
Db	1764	TTCGATACGTTGGGCGAAGGCGCGCGACGACGAGACGAAAGCAAGCGCGCGACGAGCGGAC	18233
QY	2013	GGAAGACGTAGCGGCG 2028	
Db	1824	GGAAGACGTAGCGGCG 1839	
RESULT 9			
AAH74882			
ID	AAH74882	standard; DNA; 1833 BP.	
XX	XX		
AC	AAH74882;		
XX	XX		
DT	29-OCT-2001	(first entry)	
XX	XX		
DE	Nucleotide sequence of altered recombinase ICI.		
XX	Recombinase; genetic engineering; monogenic disorder; ADA deficiency;		
KW	cystic fibrosis; familial hypercholesterolaemia; haemophilia;		
KW	chronic granulomatous disease; Duchenne's muscular dystrophy;		
KW	Fanconi's disease; anemia; sickle-cell anemia; Gaucher's disease;		
KW	Hunter's syndrome; X-linked severe combined immunodeficiency; SCID;		
KW	infectious disease; acquired disorder; tumour; cancer; ss.		
XX	XX		
OS	Synthetic.		
XX	Bacteriophage.		
XX	XX		
EH	Key	Location/Qualifiers	
FT	CDS	1..1833	
FT		/tag= a	
XX		/product= "altered recombinase ICI"	
XX	MO200161049-A1.		
XX	23-AUG-2001.		
ID	ID		





QY 1713 GCTCCGAGCAGAGGGGGGAGAGACGGCTTGCCGAACTTGAAAGCCCGGAGCCCGAA 1772  
 Db 1524 GCTCCGAGCAGAGGGGGGAGAGACGGCTTGCCGAACTTGAAAGCCCGGAGCCCGAA 1583  
 QY 1773 GCTTCCCTTACCAATGTTCCCGAGAGACCGGAGCTTACCCGAGCCGCTTAAGTC 1832  
 Db 1584 GCTTCCCTTACCAATGTTCCCGAGAGACCGGAGCTTACCCGAGCCGCTTAAGTC 1643  
 QY 1833 GTGTGGGGGGGGCGCTGCTAGTAGACAGACAGCGGTGTTCTCGGGCTTCTAGACAA 1892  
 Db 1644 GTGTGGGGGGGGCGCTGCTAGTAGACAGACAGCGGTGTTCTCGGGCTTCTAGACAA 1703  
 QY 1893 GATGTTGTACAGAGTGTGACTAGGGGAGGGGAGGAGAACGCCATCGAGAACGCCGC 1952  
 Db 1704 GATGTTGTACAGAGTGTGACTAGGGGAGGGGAGGAGAACGCCATCGAGAACGCCGC 1763  
 QY 1953 TTGCATCAGTGGGCGAGAGCCCGACCGAGACGAGACGAGACGAGACGAGACGAGAC 2012  
 Db 1764 TTGCATCAGTGGGCGAGAGCCCGACCGAGACGAGACGAGACGAGACGAGACGAGAC 1822  
 QY 2013 GGAAGACGTAG 2023  
 Db 1823 GGAAGACGTAG 1833

RESULT 10  
 AAD19829  
 ID AAD19829 standard; DNA; 912 BP.

AC AAD19829;  
 DT 18-DEC-2001 (first entry)  
 XX  
 DE Green fluorescent protein-1, syngFP1 gene with ST-LS1 intron sequence.  
 XX  
 KW Cestrum yellow leaf curling virus; CmYLCV; transcription;  
 XX transgenic plant; green fluorescent protein 1; GFP1; ds.  
 OS Solanum tuberosum.  
 XX  
 XX Unidentified.  
 FT Key Location/Qualifiers  
 FT Intron 278..465  
 FT /tag= a  
 FT /note= "Solanum tuberosum ST-LS1 intron"  
 PN WO200173087-A1.  
 XX  
 XX 04-OCT-2001.  
 FD  
 XX  
 PF 26-MAR-2001; 2001WO-EP03408.  
 XX  
 PR 27-MAR-2000; 2000GB-0007427.  
 PR 28-APR-2000; 2000GB-0010486.  
 PR 26-JAN-2001; 2001EP-0101802.  
 XX  
 PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 XX  
 PI Hohn T, Stavolone L, De Haan PT, Ligon HT, Kononova M;  
 XX  
 DR WPI; 2001-616524/71.  
 XX  
 PT Novel DNA sequence obtained from genome of Cestrum yellow leaf curling  
 PT virus for conferring constitutive expression of an associated desired  
 PT polynucleotide -  
 XX  
 PS Example 19; Page 66; 100pp; English.  
 XX  
 CC The invention relates to Cestrum yellow leaf curling virus (CmYLCV) novel  
 CC DNA sequences which functions as transcription promoters of an associated  
 CC polynucleotide sequence. These CmYLCV DNA molecules confers constitutive  
 CC expression of associated polynucleotide sequences. The invention also  
 CC relates to recombinant DNA sequences containing promoter sequences which

CC are used for creating transgenic plants expressing DNA of interest at all  
 CC times and in most tissues and organs. The present DNA sequence is green  
 CC fluorescent protein-1, syngFP1 gene with ST-LS1 intron sequence. Plant  
 CC optimized GFP reporter gene is used in the construction of plant  
 CC transformation vectors.  
 SQ Sequence 912 BP; 245 A; 245 C; 240 G; 182 T; 0 other;

Query Match 9.88; Score 198.6; DB 22; Length 912;  
 Best Local Similarity 95.88; Pred. No. 1e-34;  
 Matches 204; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 21 CGGGGTGATACGTAAGTTCTCTCTTACCTTGTGATATATATATATATATATCA 80  
 Db 265 CCGGAGGCGACGTAAGTTCTCTCTTACCTTGTGATATATATATATATATCA 324  
 QY 81 TTGATGTAAT 140  
 Db 325 TTGATGTAAT 384  
 QY 141 ATTGCTTTCTGTAGTTTATAGTGTATATATATATATATATATATATATATAT 200  
 Db 385 ATTGCTTTCTGTAGTTTATAGTGTATATATATATATATATATATATATATAT 444  
 QY 201 CCAAAATTTGTTGATGTCGAGTACGCGGTC 233  
 Db 445 CCAAAATTTGTTGATGTCGAGTACGCGGTC 477

RESULT 11  
 AAD19834  
 ID AAD19834 standard; DNA; 1577 BP.

AC AAD19834;  
 DT 18-DEC-2001 (first entry)  
 XX  
 DE Promoter-reporter cassette #3 to construct plant transformation vector.  
 XX  
 KW Cestrum yellow leaf curling virus; CmYLCV; transcription;  
 KW transgenic plant; nopaline synthase; NOS; green fluorescent protein-1;  
 KW GFP1; ds.  
 XX  
 XX Chimeric - Cestrum yellow leaf curling virus.  
 OS Chimeric - Agrobacterium tumefaciens.  
 OS Chimeric - Unidentified.  
 XX  
 FT Key Location/Qualifiers  
 FT Promoter 1..354  
 FT /tag= a  
 FT /note= "CmYLCV CmpC promoter fragment"  
 FT misc\_feature 380..1292  
 FT /tag= b  
 FT /note= "Corresponds to SyngFP1 gene"  
 FT terminator 1304..1577  
 FT /tag= c  
 FT /note= "Nopaline synthase (NOS) terminator"  
 PN WO200173087-A1.  
 XX  
 XX 04-OCT-2001.  
 PD  
 XX  
 PF 26-MAR-2001; 2001WO-EP03408.  
 XX  
 PR 27-MAR-2000; 2000GB-0007427.  
 PR 28-APR-2000; 2000GB-0010486.  
 PR 26-JAN-2001; 2001EP-0101802.  
 XX  
 PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 XX  
 PI Hohn T, Stavolone L, De Haan PT, Ligon HT, Kononova M;  
 XX  
 DR WPI; 2001-616524/71.

PD		04-0CT-2001.
PF	XX	
PE	XX	26-MAR-2001; 2001WO-EP03408.
PR	XX	27-MAR-2000; 2000GB-0007427.
PR	XX	28-APR-2000; 2000GB-0010466.
PR	XX	26-JAN-2001; 2001EP-0101802.
PA	XX	(SYGN ) SYNGENTA PARTICIPATIONS AG.
PI	XX	Hohn T, Stavolone L, De Haan PT, Liqon HT, Kononova M;
DR	XX	WPI; 2001-616524/71.
XX	XX	
P7	PT	Novel DNA sequence obtained from genome of Cestrum yellow leaf curling
PT	PT	virus for conferring constitutive expression of an associated desired
.XX	PS	polynucleotide
XX	XX	
XX	XX	Example 19; Page 68-69; 100pp; English.
CC	CC	The invention relates to Cestrum yellow leaf curling virus (CmLYCV) novel
CC	CC	DNA sequences which functions as transcription promoters of an associated
CC	CC	polynucleotide sequence. These CmLYCV DNA molecules confers constitutive
CC	CC	expression of associated polynucleotide sequences. The invention also
CC	CC	relates to recombinant DNA sequences containing promoter sequences which
CC	CC	are used for creating transgenic plants expressing DNA of interest at all
CC	CC	times and in most tissues and organs. The present DNA sequence is a
CC	CC	promoter-reporter cassette which contains CmLYCV cmfs promoter fragment,
CC	CC	green fluorescent protein-1 syngfp reporter gene and the
CC	CC	Agrobacterium tumefaciens nopaline synthase (NOS) terminator. Promoter-
CC	CC	reporter cassette is used to construct plant transformation vector.
XX	XX	
S0		Sequence 1618 BP; 468 A; 383 C; 400 G; 367 T; 0 other;
		Query Match            9.8%; Score 198.6; DB 22; Length 1618;
		Best Local Similarity   95.8%; Pred. No. 1.3e-34;
		Matches 204; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
OY		21 CGGGGTGATACGTAAGTTTCGCTTCACTTATATATATAAATTATCATTA 80
DB		684 CCGGAGGGCTACGTAAGTTTCGCTTCACTTATATATATAAATTATCATTA 743
OY		81 TTAGTAGTAATTAATTAATTTCAAATATTTTTTCAAATAAAGAATGACTATATACA 140
DB		744 TTAGTAGTAATTAATTAATTTCAAATATTTTTTCAAATAAAGAATGACTATATACA 803
OY		141 ATTGCTTTTCGTACTTATATAAGTGCTGAATTTAAATTTAACTTTCTAATATGA 200
DB		804 ATTGCTTTTCGTACTTATATAAGTGCTGAATTTAAATTTAACTTTCTAATATATGA 863
OY		201 CCAAAATTTGGATGTCAGGTCAGCGGGTGC 233
DB		864 CCAAAATTTGGATGTCAGGTCAGGAGAGCGC 896
		RESULT 13
		AAD19840
ID	XX	AAD19840 standard; DNA; 2943 BP.
XX	XX	
AC	XX	AAD19840;
XX	XX	
DT	XX	18-DEC-2001 (first entry)
XX	XX	
DE	XX	Ubp3(At)-syngFP1-NOS terminator expression cassette.
KW	XX	Cestrum yellow leaf curling virus; CmLYCV; transcription; maize;
KW	XX	transgenic plant; nopaline synthase; NOS; green fluorescent protein; GFP;
ds.	XX	
XX	XX	Chimeric - Arabidopsis thaliana.
OS	XX	Chimeric - Agrobacterium tumefaciens.
OS	XX	Chimeric - unidentified.
XX	XX	

ID	AA019830 standard; DNA; 2001 BP
XX	
AC	AAD19830;
XX	
DT	18-DEC-2001 (first entry)
DE	Beta-glucuronidase (GUS) reporter gene with ST-LSI intron sequence.
XX	
KM	Cestrum yellow leaf curling virus; CMVLCY; transcription;
XX	transgenic plant; beta-glucuronidase; GUS; ds.
OS	Solanum tuberosum.
OS	Unidentified.
XX	
FH	Key
FT	Intron
FT	Location/Qualifiers
FT	385..576
FT	/tag= a
FT	/note= "Solanum tuberosum ST-LSI intron"
PN	WO200173087-A1.
XX	
PD	04-OCT-2001.
XX	
PF	26-MAR-2001; 2001MO-EP03408.
XX	
PR	27-MAR-2000; 2000GB-0007427.
XX	
PR	28-APR-2000; 2000GB-0010486.
XX	
XX	26-JAN-2001; 2001EP-0101802.
PA	(SYGN ) SYNGENETIC PARTICIPATIONS AG.
XX	
PI	Hohn T, Stavojone L, De Haan PT, Lyon HT, Kononova M;
DR	WPI: 2001-616524/71.
-XX	
PT	Novel DNA sequence obtained from genome of Cestrum yellow leaf curling
PT	virus for conferring constitutive expression of an associated desired
PT	polynucleotide
XX	
PS	Example 19; Page 67-68; 100pp; English.
CC	The invention relates to Cestrum yellow leaf curling virus (CMVLCY) novel
CC	DNA sequences which functions as transcription promoters of an associated
CC	polynucleotide sequence. These CMVLCY DNA molecules confers constitutive
CC	expression of associated polynucleotide sequences. The invention also
CC	relates to recombinant DNA sequences containing promoter sequences which
CC	are used for creating transgenic plants expressing DNA of interest at all
CC	times and in most tissues and organs. The present DNA sequence is
CC	beta-glucuronidase (GUS) reporter gene with ST-LSI intron sequence.
CC	The GIG (GUS reporter gene with intron) gene is used in the construction
CC	of plant transformation vectors.
XX	
SQ	Sequence 2001 BP; 532 A; 451 C; 535 G; 483 T; 0 other;
Query Match	9.6%; Score 195; DB 22; Length 2001;
Best Local Similarity	100.0%; Pred. NO. 9e-34;
Matches 195; Conservative	0; Mismatches 0; Indels 0; Gaps 0
OY	30 TAGCAGATTTTGGCTTCACCTTGATATATATAAATTATCATTTATTAGTAGTA 89
DB	383 TAGCGAAGTTTGCCTTACCTTGATATATATAAATTATCATTTATTAGTAGTA 442
OY	90 ATTATATATTTCAAATATTTTTTCAAAAATAAGATGTAGTATATAGCAATTGCTTTT 149
DB	443 ATTTAATATTTTCAAATATTTTTTCAAAAATAAGAAATGTATATAGCAATTGCTTTT 502
OY	150 CTGTACTTTATAGCTGTATATTTTATATATACCTTTTCTATATATATGACCAAAATTT 209
DB	503 CTGTACTTTATAGCTGTATATTTTATATATACCTTTTCTATATATATGACCAAAATTT 562
OY	210 GTTATGATGCAGGA 224
DB	563 GTTATGATGCAGGA 577





GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 17:39:39 / Search time 3307.5 Seconds  
(without alignments)  
9944.984 Million cell updates/sec

Title: US-09-940-550A-9  
Perfect score: 2031  
Sequence: 1 atggcacaaggggtgtgtac.....cggaaagacgtacgagcgtag 2031

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1:	em_estb1:*
2:	em_estb2:*
3:	em_estb3:*
4:	em_estb4:*
5:	em_estb5:*
6:	em_estb6:*
7:	em_estb7:*
8:	em_estb8:*
9:	em_estb9:*
10:	em_estb10:*
11:	em_estb11:*
12:	em_estb12:*
13:	em_estb13:*
14:	em_estb14:*
15:	em_estb15:*
16:	em_estb16:*
17:	em_estb17:*
18:	em_estb18:*
19:	em_estb19:*
20:	em_estb20:*
21:	em_estb21:*
22:	em_estb22:*
23:	em_estb23:*
24:	em_estb24:*
25:	em_estb25:*
26:	em_estb26:*
27:	em_estb27:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	125.8	6.2	675	13	BM109636	BM109636 EST557172
C 2	125.8	6.2	804	14	BO512568	BO512568 EST619983
C 3	125.8	6.2	847	14	BO512569	BO512569 EST619984
C 4	60.4	3.0	621	17	CNS04POX	AL301578 Tetradon
C 5	58	2.9	928	17	CNS00DKY	AL071865 Drosophila
C 6	58	2.9	964	17	CNS017VX	AL108567 Drosophila

Result No.	Score	Match	Length	DB	ID	Description
C 7	57.4	2.8	752	17	BH585665	BH585665 BOGSL20TF
C 8	57.4	2.8	986	17	CNS0170T	AL107447 Drosophila
C 9	56.4	2.8	1101	17	CNS017SY	AL108460 Drosophila
C 10	56	2.8	780	17	CNS00T74	AL089326 Arabidops
C 11	56	2.8	1201	17	CNS01038	AL098462 Drosophila
C 12	55.8	2.7	1101	17	CNS00ZAL	AL097431 Drosophila
C 13	55.6	2.7	1190	17	CNS00ZON7	AL206908 Tetradon
C 14	55.4	2.7	1101	17	CNS003BD	AL064091 Drosophila
C 15	55.2	2.7	975	17	BH179465	BH179465 O14.P.10-
C 16	55	2.7	644	17	AQ961004	AQ961004 LERF158TF
C 17	55	2.7	720	17	CNS00XZ8	AL095522 Arabidops
C 18	55	2.7	925	17	CNS0091P	AL053013 Drosophila
C 19	54.8	2.7	859	12	BF272342	BF272342 GA_Eb001
C 20	54.2	2.7	1101	17	CNS00EVL	AL069706 Drosophila
C 21	54.2	2.7	1101	17	CNS00EVL	AL069706 Drosophila
C 22	54	2.7	920	17	CNS01UOL	AL168150 Tetradon
C 23	53.8	2.6	602	17	B28108	B28108 T2G12TRD TA
C 24	53.8	2.6	1101	17	CNS00CYH	AL060100 Drosophila
C 25	53.8	2.6	1101	17	CNS002DB	AL062360 Drosophila
C 26	53.4	2.6	454	17	B96681	B96681 T3001TR TAM
C 27	53.4	2.6	748	17	AQ286687	AQ286687 RPI11-81
C 28	53.4	2.6	876	17	CNS009G1	AL053529 Drosophila
C 29	53.2	2.6	272	14	BO596436	BO596436 PESTOab3
C 30	53.2	2.6	423	17	CNS00S62	AL087992 Arabidops
C 31	53.2	2.6	693	14	BO838839	BO838839 WHE3593_C
C 32	53.2	2.6	807	17	BH438258	BH438258 BOGLZ67TF
C 33	53.2	2.6	812	17	BH178455	BH178455 O11.J.02-
C 34	53.2	2.6	812	17	CNS07KRM	AL615412 T3 end of
C 35	53.2	2.6	1101	17	CNS016LI	AL106896 Drosophila
C 36	53	2.6	364	17	AQ583787	AQ583787 RPI1-11-4
C 37	53	2.6	1101	17	CNS00DT7	AL075293 Drosophila
C 38	52.6	2.6	1101	17	CNS00DEP	AL069493 Drosophila
C 39	52.4	2.6	673	17	CNS06UML	AL141695 T7 end of
C 40	52.4	2.6	905	17	CNS00KRX	AL077798 Drosophila
C 41	52.2	2.6	863	12	BF264892	BF264892 HV_CEA001
C 42	52.2	2.6	1101	17	CNS00EOL	AL069526 Drosophila
C 43	52	2.6	996	17	CNS00FOH	AL071063 Drosophila
C 44	52	2.6	1043	17	CNS0145P	AL103735 Drosophila
C 45	52	2.6	1292	13	BM463105	BM463105 AGENCOURT

# ALIGNMENTS

RESULT 1  
LOCUS BM109636/c 675 bp mRNA linear EST 26-NOV-2001  
DEFINITION EST557172 potato roots Solanum tuberosum cDNA clone cFR04H6 5' end,  
mRNA sequence.  
ACCESSION BM109636  
VERSION BM109636.1 GI:17070714  
KEYWORDS EST.  
SOURCE  
ORGANISM Solanum tuberosum  
potato.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids II; Solanales; Solanaceae; Solanum.  
1 (bases 1 to 675)  
van der Hoeven,R., Sun,H., Karamycheva,S.A., Tsai,J., Van Aken,S.,  
Ullrichback,T., Chiemlingo,A., Bougri,O., Buell,C.R., Ronning,C.,  
Tanksley,S. and Baker,B.  
Generation of ESTs from potato roots  
Unpublished (2001)  
CONTACT: Research Genetics, Libraries Division  
Tel: 1-800-711-6195  
Email: cdna@resgen.com  
For clone info: please contact Research Genetics, Libraries  
Division tel 1-800-711-6195, email cdna@resgen.com  
Seq primer: T3.  
FEATURES  
source  
Location/Qualifiers  
1..675  
/organism="Solanum tuberosum"  
/cultivar="Kennebec"

```

/organism="Solanum tuberosum"
/cultivar="Kennebec or Bintje"
/db_xref="taxon:4113"
/clone="STM22"
/clone_1lb="Generation of a set of potato
microarray analyses mixed potato tissues"

```

```

/organism="Solanium tuberosum"
/cultivar="Kennebec or Binjete"
/db_xref="taxon:4113"
/clone="STMH222"
/clone_lib="A collection of 5 sets of related clones"

```

```

chrone.lib = generation of a set of potato
microarray analyses mixed potato tissues"
/tissue_type="mixed tissues"
/lab_host="SOLR"
/notice="Vector: pblucscript SK(-); site.1: EcoRI; site.2:
XhoI; supplant: Combination of untreated, and Phytophthora
infestans-treated libraries of stolons, leaves, leaflets,

```







[illegible][illegible]



Oy	89	AATATATAATTTCAAATATTTTTCACAAATPAAGAAAGTAGTATTAAGCAATGCTTT	148
Dd	974	TTKMAAAMAAAAAMAAAAAMAAAAAMAAATPAMAMATTPTTTTTTTWTWMAAMWAATTAATT	1033
Oy	149	TCTGTAGTTAATAGTGCTATATTATTTAATTATTAACCTTTCTATA	195
Dd	1034	TTTTTTTTTMAATTTTTTTTTTTTTTTTTTTTAAAMA	1080

RESULT\_13

CNSOZON7/c  
LOCUS  
DEFINITION

CNSOZON7 1190 bp DNA linear GSS 14-MAY-2000  
Tetradon nigriviridis genome survey sequence PUC-Ori end of clone  
153P04 of library G from Tetradon nigriviridis, genomic survey  
sequence.

ACCESSION AL206908.1 GI:7865727  
VERSION  
KEYWORDS GSS; genome survey sequence.  
SOURCE Tetradon nigriviridis.  
ORGANISM Tetradon nigriviridis.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;  
Tetraodontidae; Tetraodon.  
1 (bases 1 to 1190)  
Roest-Crollius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher.C.,  
Bernot,A., Fizames.C., Mincker,P., Brotlier,P., Quetier.F.,  
Saurin,W. and Weissenbach,J.  
Human gene number estimate provided by genome wide analysis using  
Tetradon nigriviridis DNA sequence  
Unpublished  
2 (bases 1 to 1190)  
Roest-Crollius,H., Jalllon,O., Dasilva,C., Fizames.C., Fisher.C.,  
Bouneau,L., Billault,A., Quetier.F., Saurin,W., Bernot,A. and  
Weissenbach,J.  
Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetradon nigriviridis  
Unpublished  
3 (bases 1 to 1190)  
Genoscope.  
Direct Submission  
Submitted (12-Apr-2000)  
This sequence is a single read and was generated as part of a large  
scale clone-and sequencing project of the Tetradon nigriviridis  
genome. For more information, please take a look at  
<http://www.genoscope.cns.fr/Tetradon>.

JOURNAL TITLE  
REFERENCE AUTHORS  
JOURNAL TITLE  
COMMENT

FEATURES  
source Location/Qualifiers  
1..1190  
/organism="Tetradon nigriviridis"  
/db\_xref="taxon:99883"  
/clone="153P04"  
/clone\_lib="G"  
/note="Genoscope sequence ID : COnGIS3DH02SP1-end ;  
PUC-Ori"

-BASE COUNT 373 a 229 c 141 g 312 t 135 others  
ORIGIN

Query Match 2.7%; Score 55.6; DB 17; Length 1190;  
Best Local Similarity 48.8%; Pred. NO. 0.051;  
Matches 82; Conservative 18; Mismatched 68; Indels 0; Gaps 0;

Oy 48 TACCCTTGATATATATATTAATTAATCATTAATAGTAGTAATATATATTTCCAAATAT 107  
|| ::| |||| |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Dd 824 TAAWWMTATATTATTAATAAAMANAAMAWMAMWANTATTAATATATATAMWTTMANAAT 765  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Oy 108 TTTTTCACAAATAAAGAAGTAGTATATAGCAATGCTTTCTGTACTTTATAAGTGTG 167  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Dd 764 ATATATNMAAAAAAAAAAATATATATTTTATATATATATATATATATATATATATTA 705  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Oy 168 TATATATTAATTTAATCTTTCTCATAATATATGACCAAATTTGTGGAT 215  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

	657
D8	TATTTTAAATTGTAATAAATATTTTTTTGTATATATGAATTTATATAT 657
<hr/>	
RESULT 14	CNS003BD                  1101 bp      DNA        linear    GSS 03-JUN-1999
LOCUS	CNS003BD/c
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR0K08 of Rpci-96 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL064091
VERSION	AL064091.1 GI:4941847
KEYWORDS	GSS.
SOURCE	Drosophila melanogaster.
ORGANISM	Drosophila melanogaster. Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Epitheroidea; Drosophilidae; Drosophila. 1 (bases 1 to 1101)
REFERENCE	Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACS. For further information please see http://www.fruitfly.org The Bdgp Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named Rpci-98 and was constructed by partial ECORI digestion of Drosophila DNA provided by the BDGP from the isogenic strain YZ; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm. Location/Qualifiers 1..1101
FEATURES	/organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone="BACR0K08" /clone_1id="RPCI-98" /note="end = TET3"
BASE COUNT	395 a     120 c     103 g     334 t     149 others
ORIGIN	
Query Match	2.7%; Score 55.4; DB 17; Length 1101;
Best Local Similarity	50.3%; Pred. No. 0.056;
Matches	74; Conservative 21; Mismatches 52; Indels 0; Gaps 0;
OY	52 TTTCATATATATATATATATCATCTTATATAGTAGTATATATATTTCAATATTTT 111
DB	1006 TTTTATTTTTTTTTTTTMMWMTATATAMAWATATTAATAATTTWTMAATWRRTTWJ 947
OY	112 TTCAAATAAAAGAATGAGTAGATATAGCAAATGGTTTGCTGACTTATTAAGTGCTATA 171
DB	946 ARAKAANAANAKMWTMATTTTATATTAATTTTTTTTTTTTWTATTTATTTWAAMAAN 887
OY	172 TTTTAATTTATTAACCTTTCTAATATAT 198
DB	886 WTATWAMATATATTTATTTTATATATTT 860
RESULT 15	
LOCUS	BH179465                  975 bp      DNA        linear    GSS 29-OCT-2001
DEFINITION	014.P.10-21 SMBACL Schistosoma mansoni genomic clone 014P10.5', DNNA sequence.
ACCESSION	BH179465
VERSION	BH179465.1 GI:16280351



Oy	213	GAATGCGAGTACACCGGGTGGTTTACGACCCGATCGCGGAGCGAGCAAAATTCGAGCGC	272
Db	254	GGTGGACACGTACCCGGGTGGTTACGACCCGTCATCGCGAGCGGAGAAATTCGAGCGC	313
Oy	273	AGCAAGCCACACGACACAGCGCTACGGCAACGAAGACAAGCGCGCCACCTTCACGCGCA	332
Db	314	AGCAAGCCACACGACACAGCGCTACGGCAACGAAGACAAGCGCGCCACCTTCACGCGCA	373
Oy	333	AGTCGAGCGCAGCGGGGGCGGTTCAAGTTCTGTGGGCAATTCAGCGAAGCGCGCGAC	392
Db	374	AGTCGAGCGCAGCGGGGGCGGTTCAAGTTCTGTGGGCAATTCAGCGAAGCGCGCGAC	433
- Oy	393	GTGCGCGTTTGGGAGCGGGGGAGCGCCGCGAATTCGAGCGCATCCTGAACAAATGCGCGC	452
Db	434	GTGCGCGTTTGGGAGCGGGGGAGCGCCGCGAATTCGAGCGCATCCTGAACAAATGCGCGC	493
Oy	453	CGGGCGCTCAACATGATCATTTGTCTATGACGTGCGCGCTTCTGCGCGCTGAAGTCAAT	512
Db	494	CGGGCGCTCAACATGATCATTTGTCTATGACGTGCGCGCTTCTGCGCGCTGAAGTCAAT	553
Oy	513	GGAGCGCATTCGGAATGCTGTGGAATGCTGCGCTGGGGGTACAGATTTTTCACACTCA	572
Db	554	GGAGCGCATTCGGAATGCTGTGGAATGCTGCGCTGGGGGTACAGATTTTTCACACTCA	613
Oy	573	GGAAAGCGCTTTCGCGCAGGGAACGTCATGAGACCTGATTCACTGATTTTGGCGCTCGA	632
Db	614	GGAAAGCGCTTTCGCGCAGGGAACGTCATGAGACCTGATTCACTGATTTTGGCGCTCGA	673
Oy	633	CGCGTCGACAAAGAATCTTGTGCGAAGTGGCGCAAAATCTCGACACGAAGAACCCTTCA	692
Db	674	CGCGTCGACAAAGAATCTTGTGCGAAGTGGCGCAAAATCTCGACACGAAGAACCCTTCA	733

QY	17/3	GCATCCCTTGGACCAATGGTTCCCGGAAAGCGCGAGCTGACCCGACCGCCCTAATGC	1832
Db	1814	GCTTCCCTTGGACCAATGGTTCCCGGAAAGCGCGAGCGTGCACCCGACCGCCCTAATGC	1873
QY	1833	GTGCGTGGGGGGGGCGGCTCACTAGACGACAAAGCCGTTCTGTCGGGCTCTTCTAGACAA	1892
Db	1874	GTGCGTGGGGGGGGCGGCTCACTAGACGACAAAGCCGTTCTGTCGGGCTCTTCTAGACAA	1933
QY	1893	GATCGTGTTCACGAATGCACTCGACGCGGAGGGGAGGGAGAGCCCATGAGAAACCGCGC	1952
Db	1934	GATCGTGTTCACGAATGCACTCGACGCGGAGGGGAGGGAGAGCCCATGAGAAAGCGCGC	1993
QY	1953	TTTCGATCAGCTGGGGCGAAGCCCGCCGACGACGACGACGAGAAAGAGACGCCCGAGGACG	2012
Db	1994	TTTCGATCAGCTGGGGCGAAGCCCGCCGACGACGACGACGAGAAAGAGACGCCCGAGGACG	2053
QY	2013	GGAAAGCGTAAAGCGGGCTAG	2031
Db	2054	GGAAAGCGTAAAGCGGGCTAG	2072

RESULT 2

5190871-1

Patent No. 5190871

APPLICANT: COX, KAREN L.; KUHSTOSS, STUART A.; RAO, R. MANGARAJA

RICHARDSON, MARK A.; SCHONER, BRIGITTE E.; SEND, EUGENE T.

TITLE OF INVENTION: USE OF THE SITE-SPECIFIC INTEGRATING

FUNCTION OF PHAGE C31

NUMBER OF SEQUENCES: 3

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/364,959

FILING DATE: 12-JUN-1989

SEQ ID NO:1:

LENGTH: 3401

5190871-1

Query Match 89.1%; Score 1809.4; DB 6; Length 3401;

Best Local Similarity 99.7%; Pred. No. 0;

Matches 1813; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY	213	GATGTGAGGTAGACGGGGTGTCTTACGACCGCTAGTGCAGCGGAGCGCGAGAAATTGCGAGCG	272
Db	532	GGTGACACGTAAGCGGGGTGTCTTACGACCGCTAGTGCAGCGGAGCGCGAGAAATTGCGAGCG <td>591</td>	591
QY	273	AGCAAGCCCGACGACACACGCTTACGCGCCACGACGAGAAAGCGCGCGACTTTCAGCGCGA	332
Db	592	AGCAAGCCCGACGACACACGCTTACGCGCCACGACGAGAAAGCGCGCGACTTTCAGCGCGA	651
QY	333	AGTCGAGCGCGACGAGGGGCGGTTGAGGTGTGTGGGCGATTTCAGCGAAAGCGCGGCGAC	392
Db	652	AGTCGAGCGCGACGAGGGGCGGTTGAGGTGTGTGGGCGATTTCAGCGAAAGCGCGGCGAC	711
QY	393	GTGCGCGTTGCGGAGCGGCGAGCGCGCGAGTTGCGAAACGATCTCTGAACGATGCGCGC	452
Db	712	GTGCGCGTTGCGGAGCGGCGAGCGCGCGAGTTGCGAAACGATCTCTGAACGATGCGCGC	771
QY	453	CGGGCGGCTCAACATGATCATTTGTCTATGAGCTGTCCGCGCTTTCGCGCGCTGAAGTTCAT	512
Db	772	CGGGCGGCTCAACATGATCATTTGTCTATGAGCTGTCCGCGCTTTCGCGCGCTGAAGTTCAT	831
QY	513	GGAGCGCATTTCCGATTGTCTTCGGAATTGCTGCGCTGGCGGTGAGAGATTGTTCCACTCA	572
Db	832	GGAGCGCATTTCCGATTGTCTTCGGAATTGCTGCGCTGGCGGTGAGAGATTGTTCCACTCA	891
QY	573	GGAAAGCGTCTTCCGCGAGGAAAGTCATGAGACCTTATTCACCTGATTATAGCGGCTTCA	632
Db	892	GGAAAGCGTCTTCCGCGAGGAAAGTCATGAGACCTTATTCACCTGATTATAGCGGCTTCA	951
QY	633	CGGTCGCAAAAGAAATCTTCGTTGAGAGTGGGCGCAAAATTCGACACGAAAGAACTTCA	692
Db	952	CGGTCGCAAAAGAAATCTTCGTTGAGAGTGGGCGCAAAATTCGACACGAAAGAACTTCA	1011
QY	693	GGCGCAATTTGGGGGGGTGACGTGGCGGGGAAAGGCGCTTAAAGGCTTGAAGCTTGTTCGGA	752



1012 GCGCAATTGGGGGCGGATGCTGCGGGGAGAGCGCCCTTACGGCTTCGAGCTTGTTCGGA 1071  
 QY 753 GACGAGAGAGATCACGCGCCAAAGCGGCGGATGTCTCATCAACAGCTTGGCGCA 812  
 Db 1072 GACGAGAGAGATCACGCGCCAAAGCGGCGGATGTCTCATCAACAGCTTGGCGCA 1131  
 QY 813 CTCGACCACTCCCTTACCGGACCTTGTGAGTTCGAGCCCGGATTCGGTGGTG 872  
 Db 1132 CTCGACCACTCCCTTACCGGACCTTGTGAGTTCGAGCCCGGATTCGGTGGTG 1191  
 QY 873 GCGTGAATCAAGAGCGCAACACACTTCCCTTCAAGCGGCGGAGTCAAGCGGCATTC 932  
 Db 1192 GCGTGAATCAAGAGCGCAACACACTTCCCTTCAAGCGGCGGAGTCAAGCGGCATTC 1251  
 QY 933 CCGCGGCGACATCACGGGGCTTTGTAGCCGATGGAGCGCTGACGCGCTGCGAGCG 992  
 Db 1252 CCGCGGCGACATCACGGGGCTTTGTAGCCGATGGAGCGCTGACGCGCTGCGAGCG 1311  
 QY 993 CGAGACGATTGGGAAGAGACCGCTTCAAGCGCGCTGGGACCGCGGATTCGAGAT 1052  
 Db 1312 CGAGACGATTGGGAAGAGACCGCTTCAAGCGCGCTGGGACCGCGGATTCGAGAT 1371  
 QY 1053 CTTTCGAGACCGCGCTTATTTGGGGCTTGCCTGAGGTGATCTCAAGAGAGACCGG 1112  
 Db 1372 CTTTCGAGACCGCGCTTATTTGGGGCTTGCCTGAGGTGATCTCAAGAGAGACCGG 1431  
 QY 1113 CGGACGCGGACACGAGAGATTGAGGGTTACCGCATTCAGCGGACCGCATCGCTCG 1172  
 Db 1432 CGGACGCGGACACGAGAGATTGAGGGTTACCGCATTCAGCGGACCGCATCGCTCG 1491  
 QY 1173 GCGGTCGAGTTGATTCGCGGACGATCATGAGCGCGCTGATGAGTGTGAGCTTCAGG 1232  
 Db 1492 GCGGTCGAGTTGATTCGCGGACGATCATGAGCGCGCTGATGAGTGTGAGCTTCAGG 1551  
 QY 1233 GTGTTGGACGCGGAGGGGCGCGGCAAGGGGCTTCCGCGGCGCAAGCATTCCTCG 1292  
 Db 1552 GTGTTGGACGCGGAGGGGCGCGGCAAGGGGCTTCCGCGGCGCAAGCATTCCTCG 1611  
 QY 1293 CATGAGACAGCTGTACTGCGAGTGTGGCGGCTCATGACTTGAAGCGCGGAGAGATC 1352  
 Db 1612 CATGAGACAGCTGTACTGCGAGTGTGGCGGCTCATGACTTGAAGCGCGGAGAGATC 1671  
 QY 1353 GATCAAGACTCTTACCGCTGCGCGGCGGAGAGGTGTGCGACCGCTGCGACCTGAG 1412  
 Db 1672 GATCAAGACTCTTACCGCTGCGCGGCGGAGAGGTGTGCGACCGCTGCGACCTGAG 1731  
 QY 1413 GCAGGAGGACGTCGCAAGCTGAGCATGGCGGCACTGCAAGATTCGTTGCGGAACG 1472  
 Db 1732 GCAGGAGGACGTCGCAAGCTGAGCATGGCGGCACTGCAAGATTCGTTGCGGAACG 1791  
 QY 1473 CTTCAACAGATCAGGCGACCGCGGAGAGCGAGAGAGAGAGAGAGAGAGAGAG 1532  
 Db 1792 CTTCAACAGATCAGGCGACCGCGGAGAGCGAGAGAGAGAGAGAGAGAGAGAG 1851  
 QY 1533 CGCCGAGAGCTTGGGCAAGCTCAGTGAAGCGGCTGTGAGAAAGCGGCGGAACCT 1592  
 Db 1852 CGCCGAGAGCTTGGGCAAGCTCAGTGAAGCGGCTGTGAGAAAGCGGCGGAACCT 1911  
 QY 1593 TGTTCGAGAGCGCGCGGAGAGCGGCTTGAAGAGCTGTGAGAGAGAGAGAGAG 1652  
 Db 1912 TGTTCGAGAGCGCGCGGAGAGCGGCTTGAAGAGCTGTGAGAGAGAGAGAGAG 1971  
 QY 1653 AGGCGGCTGACAG 1712  
 Db 1972 AGGCGGCTGACAG 2031  
 QY 1713 GCTCGGAG 1772  
 Db 2032 GCTCGGAG 2091  
 QY 1773 GCTTCGAG 1832

Db 2092 GCTTCCCTTGAACAATGATGTTCCCGGAGAGACCGGACGCTGACCGGCGCTTAAGTC 2151  
 QY 1833 GTGTTGGGGGGCGCGCTGATGAGACAGCAAGCGCGCTTCTCGGCGCTTCTGTAACAA 1892  
 Db 2152 GTGTTGGGGGGCGCGCTGATGAGACAGCAAGCGCGCTTCTCGGCGCTTCTGTAACAA 2211  
 QY 1893 GATCGTTGTCAGAGATGACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1952  
 Db 2212 GATCGTTGTCAGAGATGACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2271  
 QY 1953 TTCGATCAGTGGGGGAGAGCGCGGACGAGCAGCAGAGAGAGAGAGAGAGAGAGAG 2012  
 Db 2272 TTCGATCAGTGGGGGAGAGCGCGGACGAGCAGCAGAGAGAGAGAGAGAGAGAGAG 2331  
 QY 2013 GGAAGACGTAGCGGCTAG 2031  
 Db 2332 GGAAGACGTAGCGGCTAG 2350

RESULT 3  
 US-08-452-267-2  
 ; Sequence 2, Application US/08452267  
 ; Patent No. 5801027

GENERAL INFORMATION:  
 APPLICANT: Bennett, Malcolm  
 APPLICANT: May, Sean  
 APPLICANT: Ramsay, Nichola  
 TITLE OF INVENTION: Control of Genes in Transgenic Plants  
 NUMBER OF SEQUENCES: 3  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Barnes & Thornbury  
 STREET: 11 South Meridian  
 CITY: Indianapolis  
 STATE: IN  
 COUNTRY: USA

ZIP: 46204  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/452,267  
 FILING DATE: 26-MAY-1995  
 CLASSIFICATION: A35

ATTORNEY/AGENT INFORMATION:  
 NAME: Breen, John P.  
 REGISTRATION NUMBER: 38,833  
 REFERENCE/DOCKET NUMBER: 6653-25744  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (317) 231-7745  
 TELEFAX: (317) 231-7433

INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2633 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 IMMEDIATE SOURCE:  
 CLONE: pUMIGIT(insert)  
 US-08-452-267-2

Query Match 9.6%; Score 195; DB 1; Length 2633;  
 Best Local Similarity 100.0%; Pred. No. 3,6e-38;  
 Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 TACGTAAGTTTCTGCTTACCTTGTATATATATATATATATATATATATATATATATAT 89  
 Db 693 TACGTAAGTTTCTGCTTACCTTGTATATATATATATATATATATATATATATATATAT 752



```

1 APPLICANT: Kriz, Alan L.
2 APPLICANT: Spencer, T. Michael
3 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE
4 TITLE OF INVENTION: IDENTIFICATION
5 NUMBER OF SEQUENCES: 19
6 CORRESPONDENCE ADDRESS:
7 ADDRESSEE: Arnold White & Durkee
8 STREET: P.O. Box 4433
9 CITY: Houston
10 STATE: TX
11 COUNTRY: USA
12 ZIP: 77210-4433
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15 COMPUTER: IBM PC compatible
16 OPERATING SYSTEM: PC-DOS/MS-DOS
17 SOFTWARE: PatentIn Release #1.0, Version #1.30
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/09/080,625
20 FILING DATE:
21 CLASSIFICATION:
22 ATTORNEY/AGENT INFORMATION:
23 NAME: Hanson, Robert E.
24 REGISTRATION NUMBER: P-42,628
25 REFERENCE/DOCKET NUMBER: DEKM:161
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE: (512) 418-3000
28 TELEFAX: (512) 474-7577
29 INFORMATION FOR SEQ ID NO: 1:
30 SEQUENCE CHARACTERISTICS:
31 LENGTH: 1701 base pairs
32 TYPE: nucleic acid
33 STRANDEDNESS: double
34 TOPOLOGY: linear
35 US-09-080-625-1
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
9
```



```

QY      23 GGGTGGAGACGTAAGCTTTCTGCTTCCTACCTTGATATATATTAAATAATATCATTAAT 82
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1982 GAGCGAAGAAGGTAACTTCTGCTTCCTACCTTGATATATATTAAATAATATCATTAAT 2041
QY      83 AGTACTAATATATATATTTCCAATATTTTTTCCAAAATAAAGAAGTAGTATATAGCAAT 142
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      2042 AGTAGTATATATATATTTCCAATATTTTTTCCAAAATAAAGAAGTAGTATATAGCAAT 2101
QY      143 TGCCTTTCTGTAGTTTATTAAGTGTGTATATTTTAAATTATATACTTTCTCAATATATGACC 202
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      2102 TGCTTTCTGTAGTTTATTAAGTGTGTATATTTTAAATTATATACTTTCTCAATATATGACC 2161
QY      203 AAAATTGTTGATGTGCAGGTA 224
        |||||
Db      2162 AAAATTGTTGATGTGCAGGTA 2183
        |||||

RESULT 12
US-09-695-782-4
: Sequence 4, Application US/09695782
: Patent No. 6433252
: GENERAL INFORMATION:
: APPLICANT: KRIZ, ALAN L.
: APPLICANT: GRIFFOR, MATTHEW
: TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE IDENTIFICATION
: FILE REFERENCE: DEMN:161USCI
: CURRENT APPLICATION NUMBER: US/09/695,782
: CURRENT FILING DATE: 2000-10-24
: PRIOR APPLICATION NUMBER: 09/080,625
: PRIOR FILING DATE: 1998-05-18
: NUMBER OF SEQ ID NOS: 19
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 4
: LENGTH: 3877
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Synthetic

```

```

US-09-695-782-4
; OTHER INFORMATION: Primer
Query Match          9.6%, Score 194; DB 4; Length 3877;
Best Local Similarity 97.5%; Pred. No. 7.6e-38;
Matches 197; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy      23  GGGTGCATACGTAAGTCTCGTCTTACCCTTGCATATATATATATATATATATAT 82
Db      1982 GAGCTGAAGGTAAGTATCTCGCTCTTACCCTTGCATATATATATATATATATAT 2041
Oy      83  AGTACTAATATATATATTTTCAAAATATTTTTCACAAATATAAGATGTATATGCAAT 142
Db      2042 AGTACTAATATATATATTTTCAAAATATTTTTCACAAATATAAGATGTATATGCAAT 2101
Oy      143 TGCCTTTCTGATGTTATTAAGTGTGTAATATTTTAATTTATTAACCTTCTTAATATATGACC 202
Db      2102 TGCCTTTCTGATGTTATTAAGTGTGTAATATTTTAATTTATTAACCTTCTTAATATATGACC 2161
Oy      203 AAAATTTGTGATGTGCAGGTA 224
Db      2162 AAAATTTGTGATGTGCAGGTA 2183

RESULT 13
US-08-817-188-5/c
; Sequence 5, Application US/08817188
; Patent No. 6074876
; GENERAL INFORMATION:
; APPLICANT: DE BLOCK, MARC
; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR
; FILE REFERENCE: 2121-0127P
; CURRENT APPLICATION NUMBER: US/08/817,188
; EARLIER FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: PCT/EP96/03366
; EARLIER FILING DATE: 1996-07-31
; EARLIER APPLICATION NUMBER: EP 9540184.6
; EARLIER FILING DATE: 1995-08-04
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 5560
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
; FEATURE:
; OTHER INFORMATION: plasmid pTHW142
; NAME/KEY: misc.feature
; LOCATION: (1)..(23)
; OTHER INFORMATION: RB: right border sequence of octopine TL-DNA from
; OTHER INFORMATION: pTI1653
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (84)..(296)
; OTHER INFORMATION: 3' g7: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium
; OTHER INFORMATION: T-DNA
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (318)..(869)
; OTHER INFORMATION: bar: region coding for phosphinotricin
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (830)..(2760)
; OTHER INFORMATION: pSSU: promoter region of Rubisco small subunit
; OTHER INFORMATION: gene of Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (2765)..(3058)
; OTHER INFORMATION: 3' untranslated region of the CamV 35S transcript
; OTHER INFORMATION: containing polyadenylation signals

```

```
Query Match          9.48; Score 190; DB 4; Length 1034;
Best Local Similarity 100.08; Pred. No. 4e-37;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0
```

Db 415 GATGTCAGG 424

RESULT 15  
US-09-249-585A-4/c  
; Sequence 4, Application US/09249585A

```

? APPLICANT: HORLICK, Robert
? TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOME
? FILE REFERENCE: 0867/0D905
? CURRENT APPLICATION NUMBER: US/09/249,585A
? CURRENT FILING DATE: 1999-02-11
? NUMBER OF SEQ ID NOS: 18
? SOFTWARE: PatentIn version 3.0
? SEQ ID NO 4
? LENGTH: 1926
? TYPE: DNA
? ORGANISM: Epstein Barr Virus
? FEATURE:

```

```

; LOCATION: (1)..(1926)
; OTHER INFORMATION: template strand of EBNA-1 DNA
HS-06-249-585A-A

```

Query Match	2.68;	Score 52.2;	DB 4;	Length 1926;
Best Local Similarity	45.0%;	Pred. No. 0.0022;		
Matches 238;	Conservative 0;	Mismatches 288;	Indels 3;	Gaps 1

QY	1495	GAGGCGCAGCAAGACAGTTGGCGCTTCTGTGGGAAGCCGCCGACGCTTGCGCAAGCTC	155
Db	799	GACGGGGAGAGACAGACCGGGGAGCAGCGGGAGAGCAGAGACGGGGAGGACGGGGAGGAC	740
QY	1555	ACTCAGAGCGCCTGTAGAGAAAGCCGGCCGACCGGGCGAACCTTGTTCGGAGCGGCCGACGCC	1614
Db	739	GAGGACGGGGAGGACGAGCAGGGGGAGCAGAGACGGGGAGGACGAGGACGGGGAGGAC	680
QY	1615	CTGAAGCGCCCTTGAAGAGCTGTATCAAGAAAGCCCGCGGCGATAGCGAGCCCGCTT	1674
Db	679	GGGGAGGACGGGGAGGAGACGAGGACCGGGGAGGAGAGAGACGGGAGGACGGGGAGGAC	620
QY	1675	GCGAGGAACACTTCGGAAGACACAGGAGCGGCTGACGCTCCGCGACGAGGGGCGAA	1734

```

Db 619 GGGAGAGACGAGGACGGGAGGAGACGAGACGGGGAGGACGGGGAGGACGGGGAGAG 560
Oy 1735 GAGCGGCTTGGCCGAACCTT---GAGCGCGCCGAAGCCCGAACCTTCCCTTGACCATGG 1791
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 559 GACGAGAGACGGGAGGAGGAGGAGACGAGACGAGGAGGAGGAGGAGGAGGAGGAGGAG 500
Oy 1792 TTCCCGAAGACGCGGACGCTGACCCGACCGGCCCTAAGTCGTGGTGGGGCGCGCTCA 1851
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 499 GACGGGAGAGACGGGAGAGAGACGAGACGGGGAGAGGAGGAGGAGGAGGAGGAGGAG 440
Oy 1852 GTAGACGACAGCGCGTCTTCGGGCTCTTCGTAGACAGATGCTTGTCAAGAGTCA 1911
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 439 GAGGAGAGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 380
Oy 1912 ACTACGGGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1971
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 379 GACGAGAGACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 320
Oy 1972 CCGCCGACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2020
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 319 GAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 271

```

Search completed: May 11, 2003, 23:19:03  
 Job time : 144 secs







Db 181 ATAACTTTCTAATATATGACCAAAATTTGATGTGACAGTACCGGGTCTGTACGAC 240  
QY 241 CGTCAGTCGCGGAGCGGAGAAATTCAGACGCGACAGCCCGACACAGCGTACGCGC 300  
Db 241 CGTCAGTCGCGGAGCGGAGAAATTCAGACGCGACAGCCCGACACAGCGTACGCGC 300  
QY 301 AACGAAGAACAGCGCGGAGCTTACGCGCGAAGTCAGCGCGAGCGGCGCGGTTACG 360  
Db 301 AACGAAGAACAGCGCGGAGCTTACGCGCGAAGTCAGCGCGAGCGGCGCGGTTACG 360  
QY 361 TTGCTGCGGCGATTCAGCGAAGCGCGGCGACGTCGCGGTTGCGGACGCGCGCGCCG 420  
Db 361 TTGCTGCGGCGATTCAGCGAAGCGCGGCGACGTCGCGGTTGCGGACGCGCGCGCCG 420  
QY 421 GAGTTCGAACGCGATTCCTAAGCAATGCGCGCGCGCGCTCAACATGATCTGTCTAT 480  
Db 421 GAGTTCGAACGCGATTCCTAAGCAATGCGCGCGCGCGCTCAACATGATCTGTCTAT 480  
QY 481 GACGTGTCGCGCTTCGCGCGCTGAAGGTGATGACGCGGATTCGCGATTTCTCGGAATTG 540  
Db 481 GACGTGTCGCGCTTCGCGCGCTGAAGGTGATGACGCGGATTCGCGATTTCTCGGAATTG 540  
QY 541 CTCGCGCTGCGCGCTGACGATTTGTTTCACCTCAGGAAGCGCTTCGCGGACGGAACGTC 600  
Db 541 CTCGCGCTGCGCGCTGACGATTTGTTTCACCTCAGGAAGCGCTTCGCGGACGGAACGTC 600  
QY 601 ATGGACCGATTCACCTGATTAATGCGCGCTGACGCGCTGCGACAAAGATCTTCGCTGAG 660  
Db 601 ATGGACCGATTCACCTGATTAATGCGCGCTGACGCGCTGCGACAAAGATCTTCGCTGAG 660  
QY 661 TCGGCGAAGATTCGACACAGCAAGAACCTTCAGCGCGAATTTGGCGGGTACGTCGCGGG 720  
Db 661 TCGGCGAAGATTCGACACAGCAAGAACCTTCAGCGCGAATTTGGCGGGTACGTCGCGGG 720  
QY 721 AAGCGCGCTTCAGCGCTTCGAGCTTTGTTTCGAGACGAAGAGATCAGCGCGCAACGCGCA 780  
Db 721 AAGCGCGCTTCAGCGCTTCGAGCTTTGTTTCGAGACGAAGAGATCAGCGCGCAACGCGCA 780  
QY 781 ATGTCATATGTCGATCAACAAGCTTGGCGACCTCGACGCACTCCCTTACCGGACCTTC 840  
Db 781 ATGTCATATGTCGATCAACAAGCTTGGCGACCTCGACGCACTCCCTTACCGGACCTTC 840  
QY 841 GAGTTCGAGCGCGACGTAATCCGGTGTGTGCGGTGAGATCAAGACGCAACAACCTT 900  
Db 841 GAGTTCGAGCGCGACGTAATCCGGTGTGTGCGGTGAGATCAAGACGCAACAACCTT 900  
QY 901 CCCCTCAAGCGCGGAGTCAAGCGCGGCGCATTCACCGCGGACGATCAAGCGCGCTTGAAG 960  
Db 901 CCCCTCAAGCGCGGAGTCAAGCGCGGCGCATTCACCGCGGACGATCAAGCGCGCTTGAAG 960  
QY 961 CGCATGAGCGCTGACGCGGCGCGGCGGAGACGATTTGGGAAGAACCGCTTCA 1020  
Db 961 CGCATGAGCGCTGACGCGGCGCGGCGGAGACGATTTGGGAAGAACCGCTTCA 1020  
QY 1021 AAGCGCTGGGACCGCGGACCGCTTATGCGAATCTTTCGCGGACCGGATTTGCGGGCTTC 1080  
Db 1021 AAGCGCTGGGACCGCGGACCGCTTATGCGAATCTTTCGCGGACCGGATTTGCGGGCTTC 1080  
QY 1081 GCGCGTGAAGTATCTACAAGAAAGAGCGGACGCGGACCGGACGCAAGATTTAGGCT 1140  
Db 1081 GCGCGTGAAGTATCTACAAGAAAGAGCGGACGCGGACCGGACGCAAGATTTAGGCT 1140  
QY 1141 TACCGCATTCACGCGGACCGGATTCAGCTCCGCGCGGTGAGCTGATTTGCGGACGATC 1200  
Db 1141 TACCGCATTCACGCGGACCGGATTCAGCTCCGCGCGGTGAGCTGATTTGCGGACGATC 1200  
QY 1201 ATCGAGCCCGGCTGAGTGTATGAGCTTACGCGGTGTGAGCGGAGCGGCGCGCAAG 1260  
Db 1201 ATCGAGCCCGGCTGAGTGTATGAGCTTACGCGGTGTGAGCGGAGCGGCGCGCAAG 1260  
QY 1261 GGGCTTTTCCGGGGGCAAGCCATTTCTGTCGCGCATGGAACGCTTACTGCGAGTGTGCG 1320  
Db 1261 GGGCTTTTCCGGGGGCAAGCCATTTCTGTCGCGCATGGAACGCTTACTGCGAGTGTGCG 1320

QY 1321 GCCGTCATGACTTCGGAAGCGCGGGGAAGAAATGCATCAAGGACTCTTACCGCTCCGCTCGC 1380  
Db 1321 GCCGTCATGACTTCGGAAGCGCGGGGAAGAAATGCATCAAGGACTCTTACCGCTCCGCTCGC 1380  
QY 1381 CGGAAGGTGTGCGACCGCGTCCGCGACCTGGGCGACGAGGAAGGCAAGCTGCAACGTCACATG 1440  
Db 1381 CGGAAGGTGTGCGACCGCGTCCGCGACCTGGGCGACGAGGAAGGCAAGCTGCAACGTCACATG 1440  
QY 1441 GCGGCACTGCAAGATTTGTTGCGGAAGCGCATCTTCAAGAAATCAGGACCGCGGAAGG 1500  
Db 1441 GCGGCACTGCAAGATTTGTTGCGGAAGCGCATCTTCAAGAAATCAGGACCGCGGAAGG 1500  
QY 1501 GACGAGAGAGCGTTGGCGCTTCCTGTGGAAGCGCGCGGACCGCTTCGCGAAGCTCACTGAG 1560  
Db 1501 GACGAGAGAGCGTTGGCGCTTCCTGTGGAAGCGCGCGGACCGCTTCGCGAAGCTCACTGAG 1560  
QY 1561 GCGCCTGAGAAAGAGCGGCGAAGCGGCGAAGCTTTGTCGAGAGCGCGGACCGCTTGAGC 1620  
Db 1561 GCGCCTGAGAAAGAGCGGCGAAGCGGCGAAGCTTTGTCGAGAGCGCGGACCGCTTGAGC 1620  
QY 1621 GCCCTTGAAGAGCTGTACGAAGACGCGCGGACGCGCTTACGAGGACCGCTTGAGC 1680  
Db 1621 GCCCTTGAAGAGCTGTACGAAGACGCGCGGACGCGCTTACGAGGACCGCTTGAGC 1680  
QY 1681 AAGCACTTCGGAAGCAAGCAGGCGAGCGCTGACGCTCCGCGACCAAGGGCGGGAAGAGCGG 1740  
Db 1681 AAGCACTTCGGAAGCAAGCAGGCGAGCGCTGACGCTCCGCGACCAAGGGCGGGAAGAGCGG 1740  
QY 1741 CTTGCCGAAGCTTGAAAGCGCGCGAAGCGCGGAAGCTTCCCTTGACCAATGTTCCCGAA 1800  
Db 1741 CTTGCCGAAGCTTGAAAGCGCGCGAAGCGCGGAAGCTTCCCTTGACCAATGTTCCCGAA 1800  
QY 1801 GACGCGAGCGTGAACCGCGCGGCGCTTAAGTCGTGTGTGGGGCGCGCTCACTAGACAC 1860  
Db 1801 GACGCGAGCGTGAACCGCGCGGCGCTTAAGTCGTGTGTGGGGCGCGCTCACTAGACAC 1860  
QY 1861 AAGCGCGTGTGTCGCGGCTTCTGTAGCAAGATCTGTTGACAGAGTTCGACTACGAGG 1920  
Db 1861 AAGCGCGTGTGTCGCGGCTTCTGTAGCAAGATCTGTTGACAGAGTTCGACTACGAGG 1920  
QY 1921 AAGGGGCAAGGAGACCGCATGAGAGCGCGCTTGTGATCAGAGTTCGACTACGAGG 1980  
Db 1921 AAGGGGCAAGGAGACCGCATGAGAGCGCGCTTGTGATCAGAGTTCGACTACGAGG 1980  
QY 1981 GACGAGCAGAAAGAGCAGCGCCAGGAGCGCAAGGAGCGTAGGGGCTAG 2031  
Db 1981 GACGAGCAGAAAGAGCAGCGCCAGGAGCGCAAGGAGCGTAGGGGCTAG 2031

RESULT 2  
US-09-940-550A-10  
Sequence 10, Application US/09940550A  
Publication No. US20030033635A1  
GENERAL INFORMATION:  
APPLICANT: MANKIN, LUKE  
TITLE OF INVENTION: SELF-EXCISING POLYNUCLEOTIDES AND USES THEREOF  
FILE REFERENCE: 16313-0055  
CURRENT APPLICATION NUMBER: US/09/940,550A  
CURRENT FILING DATE: 2001-08-27  
PRIORITY APPLICATION NUMBER: 60/227,961  
PRIORITY FILING DATE: 2000-08-25  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 10  
LENGTH: 2031  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial sequence: C3Int\* DNA  
US-09-940-550A-10

Query Match 99.2%; Score 2015; DB 9; Length 2031;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 2021; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
QY 1 ATGACACAAAGGGGTTGTCACCGGGGTGATACGTAAGTTCTGCTTACCTTGTATATA 60
DB 1 ATGACACAAAGGGGTTGTCACCGGGGTGATACGTAAGTTCTGCTTACCTTGTATATA 60
QY 61 TATATATATATATATATATATATATATATATATATATATATATATATATATATAT 120
DB 61 TATATATATATATATATATATATATATATATATATATATATATATATATATATAT 120
QY 121 AAGAATAGTATATATAGCAATGCTTTCTGTAGTTATAGTATATATATATATATAT 180
DB 121 AAGAATAGTATATATAGCAATGCTTTCTGTAGTTATAGTATATATATATATATAT 180
QY 181 ATAACTTTTCTAATATATATGACCAAAATTTGTTGATGTGACAGTACGGGGTCTTACGAC 240
DB 181 ATAACTTTTCTAATATATATGACCAAAATTTGTTGATGTGACAGTACGGGGTCTTACGAC 240
QY 241 CGTAGTCGCGCGGACCGGAGAAATTCAGCGCAGACAGCCAGACACAGCTAGCGCC 300
DB 241 CGTAGTCGCGCGGACCGGAGAAATTCAGCGCAGACAGCCAGACACAGCTAGCGCC 300
QY 301 AACGAGACAAGGGGCGGACCTTCAGCGGAGTACGAGCGGAGCGGGGCGGTTTCAGG 360
DB 301 AACGAGACAAGGGGCGGACCTTCAGCGGAGTACGAGCGGAGCGGGGCGGTTTCAGG 360
QY 361 TTGCTGCGGGCATTTTCAGCGAAGCGCGGGGACGTCGCGCTTCGGGACGCGGAGCGCCG 420
DB 361 TTGCTGCGGGCATTTTCAGCGAAGCGCGGGGACGTCGCGCTTCGGGACGCGGAGCGCCG 420
QY 421 GAGTTTCAAGCCATTCCTGAAGAAATGCGCGCGCGGGGCTCAACATATATATATAT 480
DB 421 GAGTTTCAAGCCATTCCTGAAGAAATGCGCGCGGGGCTCAACATATATATATATAT 480
QY 481 GACGTGTCGCGCTTCTCGCGCTGAGAGTATGAGACGATTCGATTCGCGAATTTG 540
DB 481 GACGTGTCGCGCTTCTCGCGCTGAGAGTATGAGACGATTCGATTCGCGAATTTG 540
QY 541 CTCGCCCTGGGCGTGCAGATTTGTTTCCACTAGGAAGGCGTCTTCGGGACGGAACGCTC 600
DB 541 CTCGCCCTGGGCGTGCAGATTTGTTTCCACTAGGAAGGCGTCTTCGGGACGGAACGCTC 600
QY 601 ATGACCTGATTCACCTGATATATGCGGCTCGACGCGTGCACAAAGATTTCTTCGTGAAG 660
DB 601 ATGACCTGATTCACCTGATATATGCGGCTCGACGCGTGCACAAAGATTTCTTCGTGAAG 660
QY 661 TCGCGAAGATTTCTGCACACGAAACCTTCAGCGCAATTTGGGCGGGTACGTGCGCGGG 720
DB 661 TCGCGAAGATTTCTGCACACGAAACCTTCAGCGCAATTTGGGCGGGTACGTGCGCGGG 720
QY 721 AAGCGCGCTTACGCGCTTCGAGCTTGTTCGAGAGGAAGAGATTCACGCGCAACGCGCGA 780
DB 721 AAGCGCGCTTACGCGCTTCGAGCTTGTTCGAGAGGAAGAGATTCACGCGCAACGCGCGA 780
QY 781 ATGCTCATATGCTCATATAACAAGCTTGCAGCTGCACACTCCCTTACCGGACCTTTC 840
DB 781 ATGCTCATATGCTCATATAACAAGCTTGCAGCTGCACACTCCCTTACCGGACCTTTC 840
QY 841 GAGTTTCAAGCCGAGCTAATCCGCTGTGTGCGTGCATCAAGACACACAAACACTT 900
DB 841 GAGTTTCAAGCCGAGCTAATCCGCTGTGTGCGTGCATCAAGACACACAAACACTT 900
QY 901 CCCTTCAAGCCGAGCTAATCCGCTGTGTGCGTGCATCAAGACACACAAACACTT 960
DB 901 CCCTTCAAGCCGAGCTAATCCGCTGTGTGCGTGCATCAAGACACACAAACACTT 960
QY 961 CGCATGACGCTGACGCGCTCCGACCGCGGGCGAGAGATTTGGAAACACCGCTTCA 1020
DB 961 CGCATGACGCTGACGCGCTCCGACCGCGGGCGAGAGATTTGGAAACACCGCTTCA 1020
```

```
QY 1021 AGCGCTGGAGACCGGCAACCGTTATGCGAATCTTCGGGACCGCGTATTGCGGCTTC 1080
DB 1021 AGCGCTGGAGACCGGCAACCGTTATGCGAATCTTCGGGACCGCGTATTGCGGCTTC 1080
QY 1081 GCCGCTGAGTATCTACAAAGAACCGGACGCGACCGGACCAAGATTGAGGGT 1140
DB 1081 GCCGCTGAGTATCTACAAAGAACCGGACGCGACCGGACCAAGATTGAGGGT 1140
QY 1141 TACCGCATTCAGCGGACCGCATACGCTCCGCGCGTTCGAGCTTGAATTCGGGACCGATC 1200
DB 1141 TACCGCATTCAGCGGACCGCATACGCTCCGCGCGTTCGAGCTTGAATTCGGGACCGATC 1200
QY 1201 ATCAGACCGCGTGAATGATATAGCTTCAGCGTGTGTGAGACGAGGAGGCGCGCAAG 1260
DB 1201 ATCAGACCGCGTGAATGATATAGCTTCAGCGTGTGTGAGACGAGGAGGCGCGCAAG 1260
QY 1261 GGGCTTTCCGGGGGCAAGCCATTCTGTCGCCATGACAGCTGTACTGCGAGTGTGC 1320
DB 1261 GGGCTTTCCGGGGGCAAGCCATTCTGTCGCCATGACAGCTGTACTGCGAGTGTGC 1320
QY 1321 GCCGTGATGACTTCGAAACGCGGGGGAAGATCGATCAAGAGACTTACCCTGCCGTGCG 1380
DB 1321 GCCGTGATGACTTCGAAACGCGGGGGAAGATCGATCAAGAGACTTACCCTGCCGTGCG 1380
QY 1381 CGAAGGTGTGTCAGACCGCTCCGACCTGGGACGACAGAGGACAGTGCAGCATG 1440
DB 1381 CGAAGGTGTGTCAGACCGCTCCGACCTGGGACGACAGAGGACAGTGCAGCATG 1440
QY 1441 GCGGCACTGACAAAGTTGTTGCGGAACGATCTTCAACAGATCAGGACGCGCAAGGC 1500
DB 1441 GCGGCACTGACAAAGTTGTTGCGGAACGATCTTCAACAGATCAGGACGCGCAAGGC 1500
QY 1501 GACGAAGAGAGTTGGGCTTCTGTGGGAAGCGCGCGGACGCTTGGGAACTGACTAG 1560
DB 1501 GACGAAGAGAGTTGGGCTTCTGTGGGAAGCGCGCGGACGCTTGGGAACTGACTAG 1560
QY 1561 GCGCCTGAGAAGAGCGGGGAACGAGGCGGAACCTTGTTCGGGAGCGCGGACCGCTGAAAC 1620
DB 1561 GCGCCTGAGAAGAGCGGGGAACGAGGCGGAACCTTGTTCGGGAGCGCGGACCGCTGAAAC 1620
QY 1621 GCCCTTGAAGAGCTGTACGAAGACCGCGGACGAGGAGCTTACAGAGACCGCTTGGCAGG 1680
DB 1621 GCCCTTGAAGAGCTGTACGAAGACCGCGGACGAGGAGCTTACAGAGACCGCTTGGCAGG 1680
QY 1681 AAGCACTTCGGAAGCAACAGGACGCTGACGCTCCGCGGACGAAGGGGCGGAAGAGGGG 1740
DB 1681 AAGCACTTCGGAAGCAACAGGACGCTGACGCTCCGCGGACGAAGGGGCGGAAGAGGGG 1740
QY 1741 CTTGCGCAACTTGAAGCGCGGCAAGCCCGGAAGCTTCCCTTGAACAATGTTCCCGGAA 1800
DB 1741 CTTGCGCAACTTGAAGCGCGGCAAGCCCGGAAGCTTCCCTTGAACAATGTTCCCGGAA 1800
QY 1801 GACGCGACGCTGACCGGACCGGCTTAAGTGTGTGGGGGCGGCTCACTAGTAGACAC 1860
DB 1801 GACGCGACGCTGACCGGACCGGCTTAAGTGTGTGGGGGCGGCTCACTAGTAGACAC 1860
QY 1861 AAGCGCGTTCGTCGCGGCTCTTCCTAGCAAGATCTGTTCAGCAAGTCACTACGCTGCGGGC 1920
DB 1861 AAGCGCGTTCGTCGCGGCTCTTCCTAGCAAGATCTGTTCAGCAAGTCACTACGCTGCGGGC 1920
QY 1921 AAGGGGCAAGGAAACCGCATCGAAGAGCGGCTGTGATCAGTGGGCGGACCGCGGAC 1980
DB 1921 AAGGGGCAAGGAAACCGCATCGAAGAGCGGCTGTGATCAGTGGGCGGACCGCGGAC 1980
QY 1981 GACGACGACGAAGACGACGCGGACGAGGACGGAAGACGTAAGCGGCTAG 2031
DB 1981 GACGACGACGAAGACGACGCGGACGAGGACGGAAGACGTAAGCGGCTAG 2031
```

RESULT 3  
US-09-940-550A-11  
; Sequence 11, Application US/09940550A  
; Publication No. US20030033635A1

```
GENERAL INFORMATION:
APPLICANT: MANKIN, LUKE
TITLE OF INVENTION: SELF-EXCISING POLYNUCLEOTIDES AND USES THEREOF
FILE REFERENCE: 16313-0055
CURRENT APPLICATION NUMBER: US/09/940,550A
PRIOR FILING DATE: 2001-08-27, 961
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 11
LENGTH: 9901
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: pbps EM051
NAME/KEY: modified_base
LOCATION: (1984)
OTHER INFORMATION: a, t, c, g, other or unknown
NAME/KEY: modified_base
LOCATION: (1986)
OTHER INFORMATION: a, t, c, g, other or unknown
NAME/KEY: modified_base
LOCATION: (5231)
OTHER INFORMATION: a, t, c, g, other or unknown
NAME/KEY: modified_base
LOCATION: (5233)
OTHER INFORMATION: a, t, c, g, other or unknown
NAME/KEY: modified_base
LOCATION: (8478)
OTHER INFORMATION: a, t, c, g, other or unknown
NAME/KEY: modified_base
LOCATION: (8480)
OTHER INFORMATION: a, t, c, g, other or unknown
US-09-940-550A-11

Query Match      99.2%; Score 2014; DB 9; Length 9901;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2031; Conservative 0; Mismatches 0; Indels 7; Gaps 1;

QY 1 ATGGCACAGAGGGTTGTGACCGGGGTGATACGTAAGTTCTCTCTACCTTGTATATA 60
DB 4984 ATGGCACAGAGGGTTGTGACCGGGGTGATACGTAAGTTCTCTCTACCTTGTATATA 5043
QY 61 TATATATATATATATATATATATATATATATATATATATATATATATATATATAT 120
DB 5044 TATATATATATATATATATATATATATATATATATATATATATATATATATATAT 5103
QY 121 AAGGAATGATATATATATATATATATATATATATATATATATATATATATATATAT 180
DB 5104 AAGGAATGATATATATATATATATATATATATATATATATATATATATATATATAT 5163
QY 181 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 240
DB 5164 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 5223
QY 241 CGTC-----AGTCGGAGAGCGGAGAAATGCGAGCGGAGAGCGGAGCGGAGCGG 293
DB 5224 CGTCGAGTATATATATATATATATATATATATATATATATATATATATATATATAT 5283
QY 294 TAGCGGCAAGAGAGAGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 353
DB 5284 TAGCGGCAAGAGAGAGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 5343
QY 354 GTTCAGAGTTCTGCGGCGATTTAGGAGAGCGGCGGAGCGGAGCGGAGCGGAGCGG 413
DB 5344 GTTCAGAGTTCTGCGGCGATTTAGGAGAGCGGCGGAGCGGAGCGGAGCGGAGCGG 5403
QY 414 GGGCCGAGAGTTGGAAGCGATCTGAGCAATGCGGCGGCGGCGGCGGCGGCGGCGG 473
DB 5404 GGGCCGAGAGTTGGAAGCGATCTGAGCAATGCGGCGGCGGCGGCGGCGGCGGCGG 5463
```

```
QY 474 TGTCTATAGAGTGTGCGGCGTTCTCGCGCCCTGANGATCATGAGCGGATTCGATGCTC 533
DB 5464 TGTCTATAGAGTGTGCGGCGTTCTCGCGCCCTGANGATCATGAGCGGATTCGATGCTC 5523
QY 534 GGAATTCCTGCGGCGGCGGAGATTTGTTTCCACTCAGAGAGGCGTCTCCGGAGGG 593
DB 5524 GGAATTCCTGCGGCGGCGGAGATTTGTTTCCACTCAGAGAGGCGTCTCCGGAGGG 5583
QY 594 AAACGTCATGAGACCTGATTCACCTGATTTATGCGGCTCGAGCGGCTCGCAAAAGATCTTC 653
DB 5584 AAACGTCATGAGACCTGATTCACCTGATTTATGCGGCTCGAGCGGCTCGCAAAAGATCTTC 5643
QY 654 GCTGAAGTCGCGGAGATTCGAGACGAGAAAGACCTTCAGCGGAGATTTGGCGGAGCT 713
DB 5644 GCTGAAGTCGCGGAGATTCGAGACGAGAAAGACCTTCAGCGGAGATTTGGCGGAGCT 5703
QY 714 CGGCGGGAAGGCGCTTACGCGCTTCGAGCTTGTTCGAGAGAGAGATCAGCGGCA 773
DB 5704 CGGCGGGAAGGCGCTTACGCGCTTCGAGCTTGTTCGAGAGAGAGATCAGCGGCA 5763
QY 774 CGGCGGGAATGCTCATGCTGCTATCAACAAGCTTGCAGCTGACCTGACCTTACCGG 833
DB 5764 CGGCGGGAATGCTCATGCTGCTATCAACAAGCTTGCAGCTGACCTGACCTTACCGG 5823
QY 834 ACCCTTCAGAGTTGAGACCGCGAGCTAATCCGCTGCTGCTGCTGAGATCAAGCGCA 893
DB 5824 ACCCTTCAGAGTTGAGACCGCGAGCTAATCCGCTGCTGCTGCTGAGATCAAGCGCA 5883
QY 894 ACACCTTCCTTCACAGCGGCGGAGTCACAGCCGATTCACCGGCGAGATCAGCGGCT 953
DB 5884 ACACCTTCCTTCACAGCGGCGGAGTCACAGCCGATTCACCGGCGAGATCAGCGGCT 5943
QY 954 TTGTAAGCGCATGAGAGCGGCTGAGCGGCTGCGGCGGCGGCGGAGAGATTTGGAGAAAGC 1013
DB 5944 TTGTAAGCGCATGAGAGCGGCTGAGCGGCTGCGGCGGCGGCGGAGAGATTTGGAGAAAGC 6003
QY 1014 CGCTTCAGAGCGGCTGAGAGCGGCGGAGTCACAGCCGATTCACCGGCGAGATTCGATTCG 1073
DB 6004 CGCTTCAGAGCGGCTGAGAGCGGCGGAGTCACAGCCGATTCACCGGCGAGATTCGATTCG 6063
QY 1074 GGGGCTTCGCGGCTGAGAGCTTACAGAGAGAGCGGCGGAGATTCGATTCGATTCG 1133
DB 6064 GGGGCTTCGCGGCTGAGAGCTTACAGAGAGAGCGGCGGAGATTCGATTCGATTCG 6123
QY 1134 TGAGGTTACCGCATTCAGCGGCGGAGTCACAGCCGATTCACCGGCGAGATTCGATTCG 1193
DB 6124 TGAGGTTACCGCATTCAGCGGCGGAGTCACAGCCGATTCACCGGCGAGATTCGATTCG 6183
QY 1194 ACCGATCATCGAGCGGCGGAGTCACAGCCGATTCACCGGCGAGATTCGATTCG 1253
DB 6184 ACCGATCATCGAGCGGCGGAGTCACAGCCGATTCACCGGCGAGATTCGATTCG 6243
QY 1254 CGGCAAGGCGCTTCGCGGCGGAGAGCAATCTGTCGCGGATTCAGCAAGCTGATTCG 1313
DB 6244 CGGCAAGGCGCTTCGCGGCGGAGAGCAATCTGTCGCGGATTCAGCAAGCTGATTCG 6303
QY 1314 GTGTGGCGCGCTCATGATTCGAGAGCGGCGGAGAGATTCAGCAAGCTTACCGCTG 1373
DB 6304 GTGTGGCGCGCTCATGATTCGAGAGCGGCGGAGAGATTCAGCAAGCTTACCGCTG 6363
QY 1374 CGGTGCGCGGAGAGTGTGCAACCCGCTCGGACCTGCGGCGAGCAGCAAGAGCAGTCAGAGT 1433
DB 6364 CGGTGCGCGGAGAGTGTGCAACCCGCTCGGACCTGCGGCGAGCAGCAAGAGCAGTCAGAGT 6423
QY 1434 CAGCATGCGGCGCATTCAGCAAGTTCGTTGCGGAGCGATTCGCAAGAGTCAAGAGCAGCAGC 1493
DB 6424 CAGCATGCGGCGCATTCAGCAAGTTCGTTGCGGAGCGATTCGCAAGAGTCAAGAGCAGCAGC 6483
QY 1494 CGAAGCGAGCAGAGAGAGAGTGTGCGGCTTCTGTGGAAGCGCGGAGCGCTTCGCAAGCT 1553
DB 6484 CGAAGCGAGCAGAGAGAGAGTGTGCGGCTTCTGTGGAAGCGCGGAGCGCTTCGCAAGCT 6543
```

OY	1554	CACGTAGGCGGCTTGAGAAAGACGGCGGAACCTTGTTTCGGAGACGGCGCCAGAC	161.3
Db	6544	CACGTAGGCGGCTTGAGAAAGACGGCGGAACCTTGTTTCGGAGACGGCGCCAGAC	6603
OY	1614	CCTGAAGCGCCCTTGAAAGACTGTACGAAGACCCCGCGGACGGCGTGACAGCAGCCGT	167.37
Db	6604	CCTGAAGCGCCCTTGAAAGACTGTACGAAGACCCCGCGGACGGCGTGACAGCAGCCGT	6665
OY	1674	TGCGACGAGACACTTCCGGAAAGCAACAGCAGCGCTGACGCTTCGGCAGCAAGGGCGGA	173.3
Db	6664	TGCGACGAGAGCACTTCCGGAAAGCAACAGCAGCGCTGACGCTTCGGCAGCAAGGGCGGA	672.2
OY	1734	AGACGCGCTTGCCCAACTTAAAGCGCGGACGCCGGAAGCTTCCCTTGACCAATGGTT	179.3
Db	6724	AGACGCGCTTCCCAACTTAAAGCGCGGACGCCGGAAGCTTCCCTTGACCAATGGTT	678.8
OY	1794	CCCCGAAAGACGCGACGCTACCCCGACGGGCCCTTAAGTCGTGTGGGGCGCGCTCAGT	185.5
Db	6784	CCCCGAAAGACGCGACGCTACCCCGACGGGCCCTTAAGTCGTGTGGGGCGCGCTCAGT	684.3
OY	1854	AGAGACAAAGGCGGTGTTCGTGGGGGCTTTCGTAGACAAGATCGTGTACAGCAACTGCAC	191.3
Db	6844	AGAGACAAAGGCGGTGTTCGTGGGGGCTTTCGTAGACAAGATCGTGTACAGCAACTGCAC	690.3
OY	1914	TACGGGCGAGGGGAGGGAACGGCCATCAGAGAAGCCGCTTGATACAGCTGGCGCAAGCC	197.3
Db	6904	TACGGGCGAGGGGAGGGAACGGCCATCAGAGAAGCCGCTTGATACAGCTGGCGCAAGCC	696.6
OY	1974	GCCGACCGACACGACGACAAGACACAGCCGCCAGAGACGGCAGGGAAGACGTAGCGGCGTAG	203.1
Db	6964	GCCGACCGACACGACGACAAGACACAGCCGCCAGAGACGGCAGGGAAGACGTAGCGGCGTAG	702.1

```

RESULT 4
US-09-788-297-20
: Sequence 20, Application US/09788297
: Patent No. US20020094516A1
: GENERAL INFORMATION:
: APPLICANT: Calos, Michele P.
: TITLE OF INVENTION: ALTERED RECOMBINASES FOR GENOME MODIFICATION
: FILE REFERENCE: 8400-0011
: CURRENT APPLICATION NUMBER: US/09/788, 297
: CURRENT FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 34
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 20
:
: LENGTH: 1842
:
: TYPE: DNA
: ORGANISM: Phage phiC31
US-09-788-297-20

```

Query Match	89.2%	Score 1811	DB 10	Length 1842
Best Local Similarity	99.7%	Pred. NO. 0		
Matches 1814; Conservative	0	Mismatches	5	Indels 0; Gaps 0;

0Y	213	GATGTGCAGAGTACACGGGGTCTCTTACGACCCGTCACCTCCGGAGGCGCCAAATTCGAGCG	272
Db	24	GGTGACACAGTACGGGGTGTCTTACGACCCGTCACCTCCGGAGGCGCCAAATTCGAGCGC	83
0Y	273	AGCAAGCCACAGCACACGCTAGCGCCACGAAGACAAAGCGGCGACCTTACGCGCA	332
Db	84	AGCAAGCCACAGCACACGCTAGCGCCACGAAGACAAAGCGGCGACCTTACGCGCA	143
0Y	333	AGTCAGGCGCAGCGGGGCGGTTACAGTTCTGCGGCATTTCACGAACGCGCGGCAC	392
Db	144	AGTCAGGCGCAGCGGGGCGGTTACAGTTCTGCGGCATTTCACGAACGCGCGGCAC	203
0Y	393	GTCGCGCTTTCGGGACGCGCGAGCCCGGAGTTTGAACGCATCCTGAACAATTCGCGCG	452
Db	204	GTCGCGCTTTCGGGACGCGCGAGCCCGGAGTTTGAACGCATCCTGAACAATTCGCGCG	253
0Y	453	CGGCGGCTCAACATGATCAATTGTCTATGACGTCGCGGCTTCTCGCGCCTGAAGGTAT	512

Db	264	CGGGCGGCTCAACATGATCATTTGTCATATGACGTGTGCGCTTCTGCGCCTGAAAGGTCAT	323
OY	513	GGAGCGCATTTCCGATTTGCTCTCGGAATTTGGCTTCGCGCCCTGGGCGTGAACGATTTGTTTCCACTCA	572
Db	324	GGAGCGCGATTCCGATTTGTCTCGGAATTTGGCTCGCGCCCTGGGCGTGAACGATTTGTTTCCACTCA	383
OY	573	GGAAAGCGCTCTTCCGGCAGGGAACGTCATGACCTGATTCACCTGATTTATGCGGCTCGA	632
Db	384	GGAAAGCGCTCTTCCGGCAGGGAACGTCATGACCTGATTCACCTGATTTATGCGGCTCGA	443
OY	633	CGGCTCGACAAAAGAAATCTTGCGTGAAGTGGGGAAGATTTGCGACGAGAAACCTTCA	692
Db	444	CGGCTCGACAAAAGAAATCTTGCGTGAAGTGGGGAAGATTTGCGACGAGAAACCTTCA	503
OY	693	GGCGGAATTTGGGGGGGTACCTGCGCGGGGAAGGGCCCTTACGGGCTTCGAGCTTGTTCGGA	752
Db	504	GGCGGAATTTGGGGGGGTACCTGCGCGGGGAAGGGCCCTTACGGGCTTCGAGCTTGTTCGGA	563
OY	753	GACGAGAGATCATCAGCGCAAGCGCGGCAATGGTCAATGTGTCATCAACAAGCTTGGCGCA	812
Db	564	GACGAGAGATCATCAGCGCAAGCGCGGCAATGGTCAATGTGTCATCAACAAGCTTGGCGCA	623
OY	813	CTCGACACATCCCTTACCGGAGCCCTTGCAGTTGCGAGCCCGGACGTAAATCCGATGGTG	872
Db	624	CTCGACACATCCCTTACCGGAGCCCTTGCAGTTGCGAGCCCGGACGTAAATCCGATGGTG	683
OY	873	GCGTGAGTCAAGACGCGCAACAACACCTTCCCTTCAAGCCGGGACGTCAAGCCGCATTTCA	932
Db	684	GCGTGAGTCAAGACGCGCAACAACACCTTCCCTTCAAGCCGGGACGTCAAGCCGCATTTCA	743
OY	933	CCCCGGCAGCATCACGGGGGCTTTGTAAAGCGCATGACGCTGACGCGGTTCGACCCGGGG	992
Db	744	CCCCGGCAGCATCACGGGGGCTTTGTAAAGCGCATGACGCTGACGCGGTTCGACCCGGGG	803
OY	993	CGAGCAGTATTTGGGAAGAAACCGCTTCAAGCGCTGAGGACCGGGACCGGTAATGCGCAT	1052
Db	804	CGAGCAGTATTTGGGAAGAAACCGCTTCAAGCGCTGAGGACCGGGACCGGTAATGCGCAT	863
OY	1053	CTTTTCGGGACCCCGGTATTTTCGGGGCTTTCGCGCGTGAAGGTATCTCAAGAAAGAACCCGA	1112
Db	864	CTTTTCGGGACCCCGGTATTTTCGGGGCTTTCGCGCGTGAAGGTATCTCAAGAAAGAACCCGA	923
OY	1113	CGGCACGCGCGACACGAGAATTAAGGGTTAACGCAATTCAGCCGACCCGATACGCTCGG	1172
Db	924	CGGCACGCGCGACACGAGAATTAAGGGTTAACGCAATTCAGCCGACCCGATACGCTCGG	983
OY	1173	GCCGGTCGAGCTTATTTGCGGACCGCATCATTCGAGCCCGGTGAGTGGTATGAGCTTCAGGC	1232
Db	984	GCCGGTCGAGCTTATTTGCGGACCGCATCATTCGAGCCCGGTGAGTGGTATGAGCTTCAGGC	1043
OY	1233	GTTGGTTGACGCGCAGGGGGGGCGGCAAGGGGGCTTTCCGGGGGGCAAGGCATTTCTGCGC	1292
Db	1044	GTTGGTTGACGCGCAGGGGGGGCGGCAAGGGGGCTTTCCGGGGGGCAAGGCATTTCTGCGC	1103
OY	1293	CATGAGCAAGCTGTACTGCGAGTGTGGCGCGCTCATGACTTGAAGCGCGGGGAAGAATC	1352
Db	1104	CATGAGCAAGCTGTACTGCGAGTGTGGCGCGCTCATGACTTGAAGCGCGGGGAAGAATC	1163
OY	1353	GATCAAGGACTTTTACCGGCTGCGGTGCGCGGGAAAGTGTGTCAACCCGTCCGACCTGGGCA	1412
Db	1164	GATCAAGGACTTTTACCGGCTGCGGTGCGCGGGAAAGTGTGTCAACCCGTCCGACCTGGGCA	1233
OY	1413	GCACGAAGGCGACGGAACGTCAGACATGGGGGCACTCGACCAAGTTTCGTTGGGGAACGAT	1472
Db	1224	GCACGAAGGCGACGGAACGTCAGACATGGGGGCACTCGACCAAGTTTCGTTGGGGAACGAT	1283
OY	1473	CTTCAACAAGTATCAGGCACGCCGAAGGCGACGAAGAGACGTTGGCGCTTCTGTGGGAAGC	1532
Db	1284	CTTCAACAAGTATCAGGCACGCCGAAGGCGACGAAGAGACGTTGGCGCTTCTGTGGGAAGC	1343
OY	1533	CGCCCGACGCTTGGGCAAGCTCACCTGAGGCGCTTGAAGAGAGCGCGAAGCGGCGAATCT	1592



|||||  
Db 1284 CTTCAACAAGATCAGAGCGACCGCGAAGGCGACGAGAGACGTTGGCGCTTGTGGGAGC 1343  
QY 1333 CGCGCGAGCGCTTGGGCAAGCTCACTGAGGGCGCTGAGAGAGCGCGCAACCGCGCAACT 1592  
Db 1344 CGCGCGAGCGCTTGGGCAAGCTCACTGAGGGCGCTGAGAGAGCGCGCAACCGCGCAACT 1403  
QY 1593 TGTTCGCGAGCGCGCGCGCGCTTGAAGAGCTGTACGAAGACCGCGCGC 1652  
Db 1404 TGTTCGCGAGCGCGCGCGCGCTTGAAGAGCTGTACGAAGACCGCGCGC 1463  
QY 1653 AGCGCGCTACGAGGAGCGCGCTTGGCAGGACACTTCCGGAAGCAACAGCGAGCGCTGAC 1712  
Db 1464 AGCGCGCTACGAGGAGCGCGCTTGGCAGGACACTTCCGGAAGCAACAGCGAGCGCTGAC 1523  
QY 1713 GCTTCGCGAGCAAGGCGGGAAGAGCGCGCTTGGCAGGACACTTGAAGCGCGCGCAAA 1772  
Db 1524 GCTTCGCGAGCAAGGCGGGAAGAGCGCGCTTGGCAGGACACTTGAAGCGCGCGCAAA 1583  
QY 1773 GCTTCGCGCTTGAACATGTTCCCGGAAGACGCGCGCTGACCGCGCGCTTGAAGTC 1832  
Db 1584 GCTTCGCGCTTGAACATGTTCCCGGAAGACGCGCGCTGACCGCGCGCTTGAAGTC 1643  
QY 1833 GTGCTGGGCGCGCGCTGACAGACGACGCGCTTGTCTCGGGCTCTTCTGAGACAA 1892  
Db 1644 GTGCTGGGCGCGCGCTGACAGACGACGCGCTTGTCTCGGGCTCTTCTGAGACAA 1703  
QY 1893 GATCGTTGTACGAAGTGTACTAGGGGAGGGGCGAGGAGCGCGCATCTGAGAGCGCGC 1952  
Db 1704 GATCGTTGTACGAAGTGTACTAGGGGAGGGGCGAGGAGCGCGCATCTGAGAGCGCGC 1763  
QY 1953 TTGATCATCGTGGGCGAAGCGCGCGCGCGCGAGCAGCAGCAGCAGCGCGCGCGCAG 2012  
Db 1764 TTGATCATCGTGGGCGAAGCGCGCGCGCGCGAGCAGCAGCAGCAGCGCGCGCAG 1823  
QY 2013 GGAAGACGTACGCGCGTAG 2031  
Db 1824 GCAAGACGTACGCGCGTAG 1842

RESULT 6  
US-09-788-297-27  
Sequence 27, Application US/09788297  
Patent No. US20020094516A1  
GENERAL INFORMATION:  
APPLICANT: Calos, Michele P.  
TITLE OF INVENTION: ALTERED RECOMBINASES FOR GENOME MODIFICATION  
FILE REFERENCE: 8400-0011  
CURRENT APPLICATION NUMBER: US/09/788, 297  
CURRENT FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: Patent Ver. 2.1  
SEQ ID NO 27  
LENGTH: 1839  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: altered  
US-09-788-297-27

Query Match 88.7% Score 1801.6; DB 10; Length 1839;  
Best Local Similarity 99.5% Pred. No. 0;  
Matches 1807; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 213 GATGTGAGGTACGCGGGTCTTACGACCTGCTCGCGGCGAGGCGAGATTGAGCGC 272  
Db 24 GGTGACACGTACGCGGGTCTTACGACCTGCTCGCGGCGAGGCGAGATTGAGCGC 83  
QY 273 AGCAAGCCAGCAGACAGCTAGCGCCACGAGACAGAGCGCGCGCACTTCAAGCGCA 332  
Db 84 AGCAAGCCAGCAGACAGCTAGCGCCACGAGACAGAGCGCGCGCACTTCAAGCGCA 143

QY 333 AGTCAGCGCGACGCGGCGCGCTTACAGTTCGTGCGGCACTTACAGCGAGCGCGCAC 392  
Db 144 AGTCAGCGCGACGCGGCGCGCTTACAGTTCGTGCGGCACTTACAGCGAGCGCGCAC 203  
QY 393 GTGCGCTTGGGAGCGCGCGAGCGCGCGAGTTCGAGCGCATCTGAAAGAAATGCGCGC 452  
Db 204 GTGCGCTTGGGAGCGCGCGAGCGCGCGAGTTCGAGCGCATCTGAAAGAAATGCGCGC 263  
QY 453 CGGCGGCTCAACATGATCTGTATAGCTGTGCGGCTTCTGCGGCTTCTGCGGCTTGAAGTCA 512  
Db 264 CGGCGGCTCAACATGATCTGTATAGCTGTGCGGCTTCTGCGGCTTCTGCGGCTTGAAGTCA 323  
QY 513 GGACCGCATTCGATTTGCTCGGAATTCGCGCTTGGGCGGAGCATTTGTTTCACTCA 572  
Db 324 GGACCGCATTCGATTTGCTCGGAATTCGCGCTTGGGCGGAGCATTTGTTTCACTCA 383  
QY 573 GGAAGCGCTTTCGCGAGGGAACGTCATGACCTGATTCACCTGATTAATGCGCTCGA 632  
Db 384 GGAAGCGCTTTCGCGAGGGAACGTCATGACCTGATTCACCTGATTAATGCGCTCGA 443  
QY 633 CGGTCGCAAAAGATCTTGTGCTAAGTGTGCGGCAAGATTCGCGACAGGAAACCTTCA 692  
Db 444 CGGTCGCAAAAGATCTTGTGCTAAGTGTGCGGCAAGATTCGCGACAGGAAACCTTCA 503  
QY 693 GCGCAATTTGGGCGGCTAGCTGCGGCGGGAAGGCGGCTTACGCGCTTGAAGCTTGTTCGGA 752  
Db 504 GCGCAATTTGGGCGGCTAGCTGCGGCGGGAAGGCGGCTTACGCGCTTGAAGCTTGTTCGGA 563  
QY 753 GACGAAGGAGATCACGCGCAACGCGCGAATGCTCATGCTCATCAACAAAGCTTGCGA 812  
Db 564 GACGAAGGAGATCACGCGCAACGCGCGAATGCTCATGCTCATCAACAAAGCTTGCGA 623  
QY 813 CTCGACACTCCCTTACCAGACCTTGTGAGTTCAGGCGCGCAAGTATCCGTGGTGTG 872  
Db 624 CTCGACACTCCCTTACCAGACCTTGTGAGTTCAGGCGCGCAAGTATCCGTGGTGTG 683  
QY 873 GCGTAGATCAAGAGCAGCAAAACACTTCCCTTCAAGCGCGGCAATCAAGCGCGCATCA 932  
Db 684 GCGTAGATCAAGAGCAGCAAAACACTTCCCTTCAAGCGCGGCAATCAAGCGCGCATCA 743  
QY 933 CCGGCGCAGCATCACGCGGCTTGTAAAGCGCATGAGCGCTGAGCGCGCGCGCGG 992  
Db 744 CCGGCGCAGCATCACGCGGCTTGTAAAGCGCATGAGCGCTGAGCGCGCGCGCGG 803  
QY 993 CGAGCAGATTGGGAAGAACCGCTTCAAGCGCTTGGGACCGCGCGCATATTCGCAAT 1052  
Db 804 CGAGCAGATTGGGAAGAACCGCTTCAAGCGCTTGGGACCGCGCGCATATTCGCAAT 863  
QY 1053 CTTTGGGACCGCGCTTATTTGGGCGGCTTGGCGCTGAGGATCTCAAGAAAGAACCGGA 1112  
Db 864 CTTTGGGACCGCGCTTATTTGGGCGGCTTGGCGCTGAGGATCTCAAGAAAGAACCGGA 923  
QY 1113 CGGACGCGCAGACACGAGGATTGAGGATTACCGCATTCAGCGCGACCGCATACGCTCG 1172  
Db 924 CGGACGCGCAGACACGAGGATTGAGGATTACCGCATTCAGCGCGACCGCATACGCTCG 983  
QY 1173 GCCGTCAGAGCTTATTTGCGGACCGATCATGAGCGCGCTGAGTGTATGAGCTTCAAGC 1232  
Db 984 GCCGTCAGAGCTTATTTGCGGACCGATCATGAGCGCGCTGAGTGTATGAGCTTCAAGC 1043  
QY 1233 GTGCTTGAAGCGGAGGCGCGGCAAGGCGCTTCCGCGGCGCAAGCATTTCTCCGC 1292  
Db 1044 GTGCTTGAAGCGGAGGCGCGGCAAGGCGCTTCCGCGGCGCAAGCATTTCTCCGC 1103  
QY 1293 CATGACAAAGCTTACTCGAGTGTGCGCGCTCATGACTTGAAGCGCGGGAAGAAATC 1352  
Db 1104 CATGACAAAGCTTACTCGAGTGTGCGCGCTCATGACTTGAAGCGCGGGAAGAAATC 1163  
QY 1353 GATCAAGACTTCTTACCGCTGCGCTGCGGAGAGGTGTGCAACCGCTGCGACCTGGGCA 1412  
Db 1164 GATCAAGACTTCTTACCGCTGCGCTGCGGAGAGGTGTGCAACCGCTGCGACCTGGGCA 1223



Query	213	GATGTCAGGTACCGCGGCTGTACGACCGTACGCGCGGAGCGGAGAAATTCGAGCCG	272
Query Match	87.8%	Score 1783;	DB 10; Length 1833;
Best Local Similarity	99.4%;	Pred. No. 0;	
Matches 1800;	Conservative	0;	Indels 10; Gaps 1;

Db	24	GGTGGACAGTAAAGCGGGGCTTTAGACCGTCAAGTCGGGGGAGGGCCGCAAAATTCGAGCC	83
QY	273	AGCAACCCCAAGGCACACACGCTGAGGCCAACAGAACAGGCGCCGACCTTCAGCGCGA	332
Db	84	AGTAACCCCAAGGCACACACGCTGAGGCCAACAGAACAGGCGCCGACCTTCAGCGCGA	143
QY	333	AGTCAGCCGACAGGGGGGCGGGTTCAGGTGTGCGGGGATTCAGCGAACGGCGCGGCAC	392
Db	144	AGTCAGCGCGACAGGGGGCGGGTTCAGGTGTGCGGGGATTCAGCGAACGGCGCGGCAC	203
QY	393	GTTCGCGTTTCGGGACGCGGGAGCGCCCGAGTTCGAACGCATCCTGAACGAATTCGCCGC	452
Db	204	GTTCGCGTTTCGGGAGCGGGGAAAGCCCGAGTTCGAACGCATCCTGAACGAATTCGCCGC	263
QY	453	CGGGCGGCTCAACATGATCATGTTGATAGAGTGTCCGCGCTTCGCGGCTCGAAGGTGAT	512
Db	264	CGGGCGGCTCAACATGATCATGTTGATAGAGTGTCCGCGCTTCGCGGCTCGAAGGTGAT	323
QY	513	GGACGCGATTCCGATTGTCTCGGAATTCCTCCGCTGGGGGTGACGATTTGTTCCACTCA	572
Db	324	GGACGCGATTCCGATTGTCTCGGAATTCCTCCGCTGGGGGTGACGATTTGTTCCACTCA	383
QY	573	GGAAAGCGCTTCGGGCGAGGAAACGTATGAGACCTGATTCACCTGATTTGCGGCTCGA	632
Db	384	GGAAAGCGCTTCGGGCGAGGAAACGTATGAGACCTGATTCACCTGATTTGCGGCTCGA	443
QY	633	CGCGTCGACAAAGAAATCTTGCTGGAAGTTCGGCGAATTCGAGACGCAAGAACCTTCA	692
Db	444	CGCGTCGACAAAGAAATCTTGCTGGAAGTTCGGCGAATTCGAGACGCAAGAACCTTCA	503
QY	693	GGCGCAATTTGGGGGGGGTACGTCTGGCGGGGAAGCGCCTTACGGCTTGAGCTTTGGGA	752
Db	504	GGCGCAATTTGGGGGGGGTACGTCTGGCGGGGAAGCGCCTTACGGCTTGAGCTTTGGGA	563
QY	753	GACGAAGAGATTCACGCGCCACGCGCAATAGTCAATGTGTATCAACAGCTTCGGA	812
Db	564	GACGAAGAGATTCACGCGCCACGCGCAATAGTCAATGTGTATCAACAGCTTCGGA	623
QY	813	CTTCGACCACTCCCTTACCGGACCCCTTCGAGTTCGAGCCCGACGTAATCCGATGATGTG	872
Db	624	CTTCGACCACTCCCTTACCGGACCCCTTCGAGTTCGAGCCCGACGTAATCCGATGATGTG	683
QY	873	GGGTAGATTCAAAGACGCGCAAAACCTTCCTTCACAGCGGGGAGTCAGACCGCCATTC	932
Db	684	GGGTAGATTCAAAGACGCGCAAAACCTTCCTTCACAGCGGGGAGTCAGACCGCCATTC	743
QY	933	CCCGGGGACATCAACGGGGCTTTGTAAGCATGAGACGCTGAGCGCCGACCGACCGGGG	992
Db	744	CCCGGGGACATCAACGGGGCTTTGTAAGCATGAGACGCTGAGCGCCGACCGGGG	803
QY	993	CGAAGCATTTGGGAAGAAGACCGCTTAAACGCGCTGGGACCCGGCAACCGTTATGCGAAT	1052
Db	804	CGAAGCATTTGGGAAGAAGACCGCTTAAACGCGCTGGGACCCGGCAACCGTTATGCGAAT	863
QY	1053	CGTTTCGGGACCGCGGTATTCGCGGGCTTCGCGGCTGAGGTGATCTACAAAGAACCCGGA	1112
Db	864	CGTTTCGGGACCGCGGTATTCGCGGGCTTCGCGGCTGAGGTGATCTACAAAGAACCCGGA	923
QY	1113	CGGACACCGGACACGAGAATTAAGGCTTACCGCATTCAGCCGACCGGATCACGCTCG	1172
Db	924	CGGACACCGGACACGAGAATTAAGGCTTACCGCATTCAGCCGACCGGATCACGCTCG	983
QY	1173	GGCGGTGAGCTGATTTGGGAGCGGACCGATTCAGAGCCCGCTGAGTGTGATGACCTTCAGC	1232
Db	984	GGCGGTGAGCTGATTTGGGAGCGGACCGATTCAGAGCCCGCTGAGTGTGATGACCTTCAGC	1043
QY	1233	GTGCTTGGAGCGGAGGGGCGCGGCAAGGGGCTTTCCCGGGGCAAGCATTTCTTCGCG	1292
Db	1044	GTGCTTGGAGCGGAGGGGCGCGGCAAGGGGCTTTCCCGGGGCAAGCATTTCTTCGCG	1103
QY	1293	CATGAGCAACGCTTACTCGGATGTGTGGCGCGCTCATGACTTGAAGCGCGGGGAGAAATC	1352
Db	1104	CATGAGCAACGCTTACTCGGATGTGTGGCGCGCTCATGACTTGAAGCGCGGGGAGAAATC	1163







LOCATION: (3529)

```

Query Match      2.5%   Score 50; DB 9; Length 32367;
Best Local Similarity 55.9%; Pred. No. 0.47; Indels 0; Gaps 0
Matches 95; Conservative 0; Mismatches 75; Indels 0; Gaps 0

OY 48 TACCTTGCATATATATATATATATATATATATATATATATAGTAGTAAATATTAATTCGCAATAT 107
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7045 TTCTTTAATATATATATATATATATATATATATATATATATATATATATATATATATATAT 6986

OY 108 TTTTTCAATAAATAAGAATAGTAGTATATAGCAATTGCTTTCTGTAGTTATAGTGCG 167
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6985 ATTATATATATATACATATATGTTTGTATATATATATATATATATATATATATATATATAT 6926

OY 168 TATATTTTAAATTTATACCTTTCTTAATATATGACCAAAATTTGGTGAATG 217
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6925 TATATTTTAAATTTATGTTTATATGTTTATATATATATATATATATATATATATATATAT 6876

RESULT 14
US-10-158-160A-14/c
Sequence 14, Application US/10158160A
Publication No. US20030059805A1
GENERAL INFORMATION:
APPLICANT: RAPPOLD-HOEHRBRAND, GUDDRUN
APPLICANT: RAO, ERCOLE
TITLE OF INVENTION: HUMAN GROWTH GENE AND SHORT STATURE GENE REGION
FILE REFERENCE: 108351-00004
CURRENT APPLICATION NUMBER: US/10/158,160A
CURRENT FILING DATE: 2002-08-20
PRIOR APPLICATION NUMBER: 09/147,699
PRIOR FILING DATE: 1999-06-24
PRIOR APPLICATION NUMBER: PCT/EP97/05355
PRIOR FILING DATE: 1997-09-29
PRIOR APPLICATION NUMBER: 60/027,633
PRIOR FILING DATE: 1996-10-01
PRIOR APPLICATION NUMBER: EP/97100583.0
PRIOR FILING DATE: 1997-01-16
NUMBER OF SEQ ID NOS: 55
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 14
LENGTH: 32367
TYPE: DNA
ORGANISM: Homo sapiens
US-10-158-160A-14


```





GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 15:44:18 ; Search time 5977.5 Seconds

(Without alignments)  
9888.381 Million cell updates/sec

Title: US-09-940-550A-10

Perfect score: 2031

Sequence: 1 atggcacaaggggtgtgtac.....cggagacgctagcggcgtag 2031

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.\*

1: gb\_da.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vl.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vl.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_mus.\*

34: em\_htg\_pln.\*

35: em\_htg\_rnd.\*

36: em\_htg\_syn.\*

37: em\_htg\_vit.\*

38: em\_sy.\*

39: em\_htgo\_hum.\*

40: em\_htgo\_mus.\*

41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2031	100.0	2031	6	AX394183
2	2015	99.2	2031	6	AX394182
3	1998	98.4	9901	6	AX394184
4	1796.6	88.5	2200	7	C31INTDNA
5	1796.6	88.5	5711	6	AX114861
6	1796.6	88.5	5715	12	PC1414670
7	1796.6	88.5	41489	7	BP6589
8	200.4	9.9	5967	12	EV084006
9	198.6	9.8	912	6	AX259239
10	198.6	9.8	1577	6	AX259244
11	198.6	9.8	1618	6	AX259242
12	198.6	9.8	2943	6	AX259250
13	195	9.6	2001	6	AX259240
14	195	9.6	2031	6	AX356663
15	195	9.6	2633	6	AR037156
16	195	9.6	2633	6	AR112042
17	195	9.6	2725	6	AX259245
18	195	9.6	2730	6	AX259243
19	195	9.6	4072	6	AX259251
20	195	9.6	4341	6	AX259249
21	195	9.6	5288	12	ASTDNABV
22	195	9.6	5534	6	AR037157
23	195	9.6	5534	6	AR112043
24	195	9.6	12817	6	AX356664
25	195	9.6	13274	6	AX356662
26	194	9.6	1701	6	AR174855
27	194	9.6	3336	6	AR174856
28	194	9.6	3877	6	AR174858
29	190.2	9.4	5560	6	AR098311
30	190.2	9.4	1034	6	AR87679
31	190	9.4	39314	1	SGR300302
32	181.4	8.9	363	6	AX060392
33	180.6	8.9	363	6	AX060389
34	180.6	8.9	363	6	AX060391
35	180.6	8.9	369	6	AX060390
36	179.8	8.9	3699	8	STLSIG
37	62.2	3.1	39314	1	SGR300302
38	61.2	3.0	125020	9	AR429315
39	60	3.0	125020	9	AR429315
40	59	2.9	134375	2	AC118169
41	58.6	2.9	169546	2	AC004157
42	58.2	2.9	154902	9	AC005799
43	58.2	2.9	191947	2	AC079210
44	58	2.9	9424	8	SP0404228
45	57.8	2.8	1181	8	MISC0X13

## ALIGNMENTS

RESULT 1	AX394183	2031 bp	DNA	Linear	PAT 23-MAR-2002
LOCUS	AX394183				
DEFINITION	Sequence 10 from Patent WO0216609.				
ACCESSION	AX394183				
VERSION	AX394183.1	GI:19702117			
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS	Mankin, L.				
TITLE	Self-excising polynucleotides and uses thereof				
JOURNAL	Patent: WO 0216609-A 10 28-FEB-2002;				
	Basf Plant Science GmbH (DE) ; McKersie, Bryan (US)				

Filed at US

FEATURES  
Source Location/Qualifiers  
1. 2031  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="C3int+ DNA sequence"  
BASE COUNT 454 a 558 c 640 g 379 t  
ORIGIN

Query Match 100.0%; Score 2031; DB 6; Length 2031;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2031; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCACAAGGGGTTGACCGGGGATACGTAAGTTTCTTACCTTGATATA 60  
Db 1 ATGGCACAAGGGGTTGACCGGGGATACGTAAGTTTCTTACCTTGATATA 60  
QY 61 TAT 120  
Db 61 TAT 120  
QY 121 AAAGAT 180  
Db 121 AAAGAT 180  
QY 181 ATATCTTTTCTATATATATATATATATATATATATATATATATATAT 240  
Db 181 ATATCTTTTCTATATATATATATATATATATATATATATATATATAT 240  
QY 241 CGTCAGTCGGCGAGCGGAGATAGCAGACAGCCAGCAGACAGCGTACGC 300  
Db 241 CGTCAGTCGGCGAGCGGAGATAGCAGACAGCCAGCAGACAGCGTACGC 300  
QY 301 AACGAACACAAGGGGCGGCGCTTACGCGGAGTGCAGCGCGCGGCGTTC 360  
Db 301 AACGAACACAAGGGGCGGCGCTTACGCGGAGTGCAGCGCGCGGCGTTC 360  
QY 361 TTGCTCGGGCATTTCAACGAGGCGCGGCGCGCTTCCGAGCGCGCGCGC 420  
Db 361 TTGCTCGGGCATTTCAACGAGGCGCGGCGCGCTTCCGAGCGCGCGCGC 420  
QY 421 GAGTTGCAAGCATCTTCAACGAGGCGCGGCGCGCTTCCGAGCGCGCGC 480  
Db 421 GAGTTGCAAGCATCTTCAACGAGGCGCGGCGCGCTTCCGAGCGCGCGC 480  
QY 481 GAGGTGCGGCTTTCGCGCTGGAAGGTATGAGGCGGCGGCGGCGGCGG 540  
Db 481 GAGGTGCGGCTTTCGCGCTGGAAGGTATGAGGCGGCGGCGGCGGCGG 540  
QY 541 CTGCGCCCTGGGCGTGAAGATTTTCCACTCAGAGAGGCGGCTTCCGAG 600  
Db 541 CTGCGCCCTGGGCGTGAAGATTTTCCACTCAGAGAGGCGGCTTCCGAG 600  
QY 601 ATGACCTGATTCACCTGATATGCGGCTGAGCGCTGCGACAAATCTTCC 660  
Db 601 ATGACCTGATTCACCTGATATGCGGCTGAGCGCTGCGACAAATCTTCC 660  
QY 661 TCGGCGAAGATTTCTGACAGAGAACTTACGCGGAGTTGGGCGGCTTCC 720  
Db 661 TCGGCGAAGATTTCTGACAGAGAACTTACGCGGAGTTGGGCGGCTTCC 720  
QY 721 AAGGCGCTTACGCGCTTCAAGCTTTTTCGAGACGAAAGATATCAGCG 780  
Db 721 AAGGCGCTTACGCGCTTCAAGCTTTTTCGAGACGAAAGATATCAGCG 780  
QY 781 ATGATCAATGCTGATCAACAAGTTAGGCACTGACACAGCCCTTACCG 840  
Db 781 ATGATCAATGCTGATCAACAAGTTAGGCACTGACACAGCCCTTACCG 840  
QY 841 GAGTTGAGACCGCGAGCTAATCGGCTGCTGAGATCAAGACGACAAACCT 900  
Db 841 GAGTTGAGACCGCGAGCTAATCGGCTGCTGAGATCAAGACGACAAACCT 900  
QY 901 CCTTCAAGCGGCGAGTCAAGCGCCATTCACCGGCGGAGCATCAGGGGCT 960

Db 901 CCTTCAAGCGGCGAGTCAAGCGCCATTCACCGGCGGAGCATCAGGGGCT 960  
QY 961 CGCATGACGCTGACCGCGGTGCGGACCGGGGAGACGATTTGGAGAAAGCC 1020  
Db 961 CGCATGACGCTGACCGCGGTGCGGACCGGGGAGACGATTTGGAGAAAGCC 1020  
QY 1021 AGCGCTGGGACCGGCGGAGTATGCGATCTTCCGAGACCGGCTATTTGGG 1080  
Db 1021 AGCGCTGGGACCGGCGGAGTATGCGATCTTCCGAGACCGGCTATTTGGG 1080  
QY 1081 GCGCTGAGGTGATCTACAAAGAAAGACCGGAGCGGACCGGACGAAATTT 1140  
Db 1081 GCGCTGAGGTGATCTACAAAGAAAGACCGGAGCGGACCGGACGAAATTT 1140  
QY 1141 TACCGATTTAGGCGGACCGGATCAGGCTCCGCGGCTGAGCTTGAATTC 1200  
Db 1141 TACCGATTTAGGCGGACCGGATCAGGCTCCGCGGCTGAGCTTGAATTC 1200  
QY 1201 ATGAGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1260  
Db 1201 ATGAGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1260  
QY 1261 GGGCTTTCCGGGGGCAAGCATTTCTGCGCATGAGACAGCTGATGAGTG 1320  
Db 1261 GGGCTTTCCGGGGGCAAGCATTTCTGCGCATGAGACAGCTGATGAGTG 1320  
QY 1321 GCGCTATGATCTTCAAGGCGGCGGGAAGATGATGATGATGATGATGAT 1380  
Db 1321 GCGCTATGATCTTCAAGGCGGCGGGAAGATGATGATGATGATGATGAT 1380  
QY 1381 CGGAAGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440  
Db 1381 CGGAAGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440  
QY 1441 GCGGCACTGACGATGATGATGATGATGATGATGATGATGATGATGAT 1500  
Db 1441 GCGGCACTGACGATGATGATGATGATGATGATGATGATGATGATGAT 1500  
QY 1501 GAGGAGAGACGTTGGCGCTTCTGTTGGAGAGCGCGCGCTGCGCAAGCT 1560  
Db 1501 GAGGAGAGACGTTGGCGCTTCTGTTGGAGAGCGCGCGCTGCGCAAGCT 1560  
QY 1561 GGGCTGAGAAAGCGGCGGAGCGGCGGAGACCTTTGCGAGCGCGCGCT 1620  
Db 1561 GGGCTGAGAAAGCGGCGGAGCGGCGGAGACCTTTGCGAGCGCGCGCT 1620  
QY 1621 GCGCTTGAAGAGCTGATGATGATGATGATGATGATGATGATGATGAT 1680  
Db 1621 GCGCTTGAAGAGCTGATGATGATGATGATGATGATGATGATGATGAT 1680  
QY 1681 AAGCACTTCCGGAAGCAACAGGCGGAGCGCTGCGGACGAGGCGGAGAG 1740  
Db 1681 AAGCACTTCCGGAAGCAACAGGCGGAGCGCTGCGGACGAGGCGGAGAG 1740  
QY 1741 CTGTCGAACTTGAAGCGCGGCGGAGCGGAGGCTTGGACCAATGCTTCC 1800  
Db 1741 CTGTCGAACTTGAAGCGCGGCGGAGCGGAGGCTTGGACCAATGCTTCC 1800  
QY 1801 GAGCGCGAGCGTGAAGCGGCGGCGGAGGCGGAGGCGGCGGCGGCGG 1860  
Db 1801 GAGCGCGAGCGTGAAGCGGCGGCGGAGGCGGAGGCGGCGGCGGCGG 1860  
QY 1861 AAGCGCGTGTGCTGAGCGGCGGCGGAGGCGGAGGCGGCGGCGGCGG 1920  
Db 1861 AAGCGCGTGTGCTGAGCGGCGGCGGAGGCGGAGGCGGCGGCGGCGG 1920  
QY 1921 AAGGCGGAGGAGCGGCGGAGGAGGAGGCGGCGGCGGCGGCGGCGG 1980  
Db 1921 AAGGCGGAGGAGCGGCGGAGGAGGAGGCGGCGGCGGCGGCGGCGG 1980  
QY 1981 GAGCAGCAG 2031



Db	1981	GACGACGACGAAGACGACGCCCGCAGGACGGCAGCAAGACTAGCGCGGTAG	2031
RESULT 2			
AX394182			
LOCUS	AX394182	2031 bp	DNA
DEFINITION	Sequence 9 from Patent WO0216609.		Linear
ACCESION	AX394182		PAT 23-MAR-2002
VERSION	AX394182.1		
KEYWORDS	GI:19702116		
SOURCE			
ORGANISM	synthetic construct. artificial sequences.		
REFERENCE	1		
AUTHORS	Mankin, L.		
TITLE	Self-excising polynucleotides and uses thereof		
JOURNAL	Patent: WO 0216609-A 9 28-FEB-2002;		
FEATURES	Basf Plant Science GmbH (DE) ; McKersie, Bryan (US)		
source	Location/Qualifiers 1..2031		
BASE COUNT	451 a 562 c 640 g 378 t		
ORIGIN			
Query Match	99.2% Score 2015; DB 6; Length 2031; Best Local Similarity 99.5%; Pred. No. 0;		
Matches 2021; Conservative	0; Mismatches 10; Indels 0; Gaps 0;		
QY	1 ATGGCACAAGGGGCTGTGACCGGGTGATACGTAAGTTCTTCTTACCTTGATATA	60	
DB	1 ATGGCACAAGGGGCTGTGACCGGGTGATACGTAAGTTCTTCTTACCTTGATATA	60	
QY	61 TATTATAATTAATTCATTATATAGTAGTAATTAATTAATTTCAAATATTTTTCAAATA	120	
DB	61 TATTATAATTAATTCATTATATAGTAGTAATTAATTAATTTCAAATATTTTTCAAATA	120	
QY	121 AAGAATGATGATATATAGCAATTCCTTTCTGTAGTTTAAAGTGATATTTAATTTAATTT	180	
DB	121 AAGAATGATGATATATAGCAATTCCTTTCTGTAGTTTAAAGTGATATTTAATTTAATTT	180	
QY	181 ATAATCTTTCTTAATATATGACCAAAATTTGTTGATGTGCAGTACGGGGCTTACGAC	240	
DB	181 ATAATCTTTCTTAATATATGACCAAAATTTGTTGATGTGCAGTACGGGGCTTACGAC	240	
QY	241 CGTCAGTCGGGGAGGCGGAGAAATGACATGACGACGAAGCCGACAGCGTACGGCGCC	300	
DB	241 CGTCAGTCGGGGAGGCGGAGAAATGACATGACGACGAAGCCGACAGCGTACGGCGCC	300	
QY	301 AACGAAGACAAAGCGCGCCACCTTCAGCGCGAAGTCGACGCGAGCGGGGGCCGTTACAG	360	
DB	301 AACGAAGACAAAGCGCGCCACCTTCAGCGCGAAGTCGACGCGAGCGGGGGCCGTTACAG	360	
QY	361 TTTCGTGGGCAATTCAGCGAAGCGCGCGGACGCTCGGCTTCGGGAGCGGAGCGCCCG	420	
DB	361 TTTCGTGGGCAATTCAGCGAAGCGCGCGGACGCTCGGCTTCGGGAGCGGAGCGCCCG	420	
QY	421 GAGTTTCGAAGCATCTCTGAACGAATGCGCGCGCGGGCGGCTAACATGATCTATTGTAT	480	
DB	421 GAGTTTCGAAGCATCTCTGAACGAATGCGCGCGCGGGCGGCTAACATGATCTATTGTAT	480	
QY	481 GACGTCGCGCGCTTCGCGGCGCTGAAGGTCATGAGCGGCAATTCGATGTCGCGGAATTG	540	
DB	481 GACGTCGCGCGCTTCGCGGCGCTGAAGGTCATGAGCGGCAATTCGATGTCGCGGAATTG	540	
QY	541 CTCGCGCTGGCGGTGACGATTTGTTTCACACTGAGAAAGGCGCTTCGCGGAGGAAACGTC	600	
DB	541 CTCGCGCTGGCGGTGACGATTTGTTTCACACTGAGAAAGGCGCTTCGCGGAGGAAACGTC	600	
QY	601 ATGGAACGATGATCACCTGATTTATGGGGCTGACGCGGTGCGACAAAGATCTTCGCTGAAG	660	
DB	601 ATGGAACGATGATCACCTGATTTATGGGGCTGACGCGGTGCGACAAAGATCTTCGCTGAAG	660	

QY	661	TCGCGAAGATTCCTCGACAGAGAAGAACCTTTCACCGCGAATTTGGGGGGGTACGTTCGCGGG	720
Db	661	TCGGGGAAGATTTCTCGACACAGGAAGACCTTTCACCGGGAATTTGGGGGGGTACGTTCGCGGG	720
QY	721	AAGCGGCTTTCAGGCTTTCGAGCTTGTATTGGAGACCAAGAAATATCACGGCACAGGGCCGA	780
Db	721	AAGGCGCTTTCAGGCTTTCGAGCTTGTATTGGAGACCAAGAAATATCACGGCGAACGGCCGA	780
QY	781	ATGGTCAATGTCTCATCAACAAGATTAGCGCACTGCACACTCCCTTATCCGGACCTTTC	840
Db	781	ATGGCAATGTCTCATCAACAAGATTAGCGCACTGCACACTCCCTTATCCGGACCTTTC	840
QY	841	GAGTTCCAGCCCCGAGAAATCCGGGTGGTGGCGGTGAGATCAATAACGCACAAACACTT	900
Db	841	GAGTTCCAGCCCCGAGAAATCCGGGTGGTGGCGGTGAGATCAATAAGCGACAAACACTT	900
QY	901	CCCTTCAAGCCGGGCACTCAAGCGCGCACTTTCACCCGGGAGCATCACGGGCTTTGTATAG	960
Db	901	CCCTTCAAGCCGGGCACTTCAAGCGCGCACTTTCACCCGGGAGCATCACGGGCTTTGTATAG	960
QY	961	CGCATGCACTGACGCGCTGCGGACCCGGGGCGAGACGATTTGGAGAGACGCGTTTCA	1020
Db	961	CGCATGCACTGACGCGCTGCGGACCCGGGGCGAGACGATTTGGGAGAGACGCGTTTCA	1020
QY	1021	AGCGCTGGGAGCCCGGCAACCGTTATGCGAATCTTTCGGGAGCCCGGTATTTGGGGCTTC	1080
Db	1021	AGCGCTGGGAGCCCGGCAACCGTTATGCGAATCTTTCGGGAGCCCGGTATTTGGGGCTTC	1080
QY	1081	GCCGCTGAGGATCTCTACAAGAAAGACCGGAGCGCACGCCACACAGCAAGATTGAGGGT	1140
Db	1081	GCCGCTGAGGATCTCTACAAGAAAGACCGGAGCGCACGCCACACAGCAAGATTGAGGGT	1140
QY	1141	TACCGCATTCAGCGCGACCCGATACGCTTCGGCGCTGAGCTTGATTTGGGACCGATC	1200
Db	1141	TACCGCATTCAGCGCGACCCGATACGCTTCGGCGCTGAGCTTGATTTGGGACCGATC	1200
QY	1201	ATTCAGCCCCGTGATGCTGATGAGCTTCACGCGTGGTGTGAGCGCAGGGGGCGCGCAAG	1260
Db	1201	ATTCAGCCCCGTGATGCTGATGAGCTTCACGCGTGGTGTGAGCGCAGGGGGCGCGCAAG	1260
QY	1261	GGGCTTTTCCCGGGGGCAAGCACTTTCCTCCGCATGAGCAACCTGATCTGCAAGTATGTC	1320
Db	1261	GGGCTTTTCCCGGGGGCAAGCACTTTCCTCCGCATGAGCAACCTGATCTGCAAGTATGTC	1320
QY	1321	GCCGTCATGACTTCGAAAGCGCGGGGAAGATTCATCAAGACTTTTACCCTGCCCTGC	1380
Db	1321	GCCGTCATGACTTCGAAAGCGCGGGGAAGATTCATCAAGACTTTTACCCTGCCCTGC	1380
QY	1381	CGGAAGGTGTCGACCCGCTCCGCACTCGGGGACACAGGAAGCAACGTGCAACGTACAGCATG	1440
Db	1381	CGGAAGGTGTCGACCCGCTCCGCACTCGGGGACACAGGAAGCAACGTGCAACGTACAGCATG	1440
QY	1441	GCGGCACTCGACAGTTCTGTTTGGGGAAGCGATTTCAACAAGATCAGGACCCCAAGC	1500
Db	1441	GCGGCACTCGACAGTTCTGTTTGGGGAAGCGATTTCAACAAGATCAGGACCCCAAGC	1500
QY	1501	GACGAAGAGAGCTTGGGCTTCTGTGGGAAGCGCCGACGCTTCGGCAAGCTCACTAG	1560
Db	1501	GACGAAGAGAGCTTGGGCTTCTGTGGGAAGCGCCGACGCTTCGGCAAGCTCACTAG	1560
QY	1561	GCGGCTGGAAGAGACGGCGGAACGGGGCAACCTTGTGGCAAGCGCGCACGCCCTTGAC	1620
Db	1561	GCGGCTGGAAGAGACGGCGGAACGGGGCAACCTTGTGGCAAGCGCGCACGCCCTTGAC	1620
QY	1621	GCCCTTGAAGAGCTGTACGAAGAACCGGGCGGAGAGCTTTAGAGAGGACCCGTTGGCAGG	1680
Db	1621	GCCCTTGAAGAGCTGTACGAAGAACCGGGCGGAGAGCTTTAGAGAGGACCCGTTGGCAGG	1680
QY	1681	AAGCACTTCGGAAGCAACAGGCGACGCGTACGCTTCGGGACGCAAGGGGGGGAAGAGCGG	1740
Db	1681	AAGCACTTCGGAAGCAACAGGCGACGCGTACGCTTCGGGACGCAAGGGGGGGAAGAGCGG	1740

1741 CTTCGCCAAGCTTGAAGCCGCGAAGCCCGAAGTTGCCCTTGAACCAATGTTCCCGGAA 1800  
1741 CTTCGCCAAGCTTGAAGCCGCGAAGCCCGAAGCTTCCCTTGAACCAATGTTCCCGGAA 1800  
1801 GAGCCCGGAGTGTGACCCGACCGGCGCTTAAGTCGTGTGGGGGGCGGCTCAGTAGACAGAC 1860  
1801 GAGCCCGGAGTGTGACCCGACCGGCGCTTAAGTCGTGTGGGGGGCGGCTCAGTAGACAGAC 1860  
1861 AAGCGCGTGTCTCGTGGGCTCTTCTGTAGACAGATCGTTGTACAGAGTCGACTACGAGG 1920  
1861 AAGCGCGTGTCTCGTGGGCTCTTCTGTAGACAGATCGTTGTACAGAGTCGACTACGAGG 1920  
1921 AGGGGCGAAGGAGACGCCCATTCAGAGACGCGCTTCATCAGCTGGGCGAAGCCGCGAC 1880  
1921 AGGGGCGAAGGAGACGCCCATTCAGAGACGCGCTTCATCAGCTGGGCGAAGCCGCGAC 1880  
1981 GAGCAGCAGCAGACGACGCGCCAGAGGAGCGGACGAGACGATAGGGGCGTAG 2031  
1981 GAGCAGCAGCAGACGACGCGCCAGAGGAGCGGACGAGACGATAGGGGCGTAG 2031

RESULT 3  
AX394184 9901 bp DNA linear PAT 23-MAR-2002  
LOCUS AX394184  
DEFINITION Sequence 11 from Patent WO0216609.  
ACCESSION AX394184  
VERSION AX394184.1 GI:19702118  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1  
AUTHORS Mankin, L.  
TITLE Self-excising polynucleotides and uses thereof  
JOURNAL Patent: WO 0216609-A 11 28-FEB-2002;  
BASF Plant Science GmbH (DE); McKersie, Bryan (US)  
FEATURES  
source 1..9901  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="pBPS EM051 'T-DNA region'"  
BASE COUNT 2489 a 2371 c 2478 g 2551 t. 12 others  
ORIGIN

Query Match 98.4% Score 1998; DB 6; Length 9901;  
Best Local Similarity 99.2%; Pred. No. 0;  
Matches 2021; Conservative 0; Mismatches 10; Indels 7; Gaps 1;

1 ATGGCACAAGGGGTGTGACCGGGGTGATAGCTAGTTTGTCTTCTACCTTTGATATA 60  
4984 ATGGCACAAGGGGTGTGACCGGGGTGATAGCTAGTTTGTCTTCTACCTTTGATATA 5043  
61 TAT 120  
5044 TAT 5103  
121 AAGAGATGAT 180  
5104 AAGAGATGAT 5163  
181 ATAACTTTTCTAAT 240  
5164 ATAACTTTTCTAAT 5223  
241 CGTC-----AGTGGCGGAGCGGAGATATAGAGTGCAGACGCCACGACACAGCG 293  
5224 CTTGCRCHNTNDGTCGCGGAGCGGAGATATTCAGCGCAGCAACCCACGACACAGCG 5283  
294 TAGCGCCAAGCAAGCAAGCGCGGACCTTCAGCGCGAAGTCGAGCGGAGCGGCGCGCG 353  
5284 TAGCGCCAAGCAAGCAAGCGCGGACCTTCAGCGCGAAGTCGAGCGCGGAGCGCGCG 5343  
354 GTTCAAGTTTCGTGGGCAATTTTCAGCAAGCGCGGCGACGTCGCGGCTTCGAGCGCGCA 413

5344 GTTCAGGTTGCTGGGGGATTTTCAAGCAAGCGCGGACGCTTGGGAGCGCGGA 5403  
414 GCGCCCGGAGTGTGAAAGCATCTGAAAGCAATCCCGGCGGCGGCTCAACATGATCAT 473  
5404 GCGCCCGGAGTGTGAAAGCATCTGAAAGCAATCCCGGCGGCGGCTCAACATGATCAT 5463  
474 TGTCTATAGAGTGTGCGGCGCTTCCGCGCTGAAAGGTCATGAGCGCATTCGATTGCTC 553  
5464 TGTCTATAGAGTGTGCGGCGCTTCCGCGCTGAAAGGTCATGAGCGCATTCGATTGCTC 5523  
534 GGAATTCCTCGCCCTGAGGCGTACAGATTGTTTCCACTCAGGAAGGCGTCTCCGCGAGG 593  
5524 GGAATTCCTCGCCCTGAGGCGTACAGATTGTTTCCACTCAGGAAGGCGTCTCCGCGAGG 5583  
594 AAACGTCATGAGCATCTTACCTGATTTATGCGGCTGACGCGCTGCACAAAGATCTTC 653  
5584 AAACGTCATGAGCATCTTACCTGATTTATGCGGCTGACGCGCTGCACAAAGATCTTC 5643  
654 GCTGAAGTCGCGGAGATCTCTGACAGCAAGAACCTTCAGCGGATTTGGCGGGTACGT 713  
5644 GCTGAAGTCGCGGAGATCTCTGACAGCAAGAACCTTCAGCGGATTTGGCGGGTACGT 5703  
714 CCGCGGGGAAGGCGCTTACGCGCTTCAGGCTTGTTCGAGAGCAAGAGATCAGCGCAA 773  
5704 CCGCGGGGAAGGCGCTTACGCGCTTCAGGCTTGTTCGAGAGCAAGAGATCAGCGCAA 5763  
774 CCGCGGAATGTCATGTCGTGATCAACAAGTTAGGCGCATGACACATCCCTTACCGG 833  
5764 CCGCGGAATGTCATGTCGTGATCAACAAGTTAGGCGCATGACACATCCCTTACCGG 5823  
834 ACCCTTCGAGTTGAGACCCCGAGCTAATCCGTTGTGTGGCTGTGAGATCAAGACGACA 893  
5824 ACCCTTCGAGTTGAGACCCCGAGCTAATCCGTTGTGTGGCTGTGAGATCAAGACGACA 5883  
894 AACACCTTCCTTCAAGCGCGGCAATCAGCCGCTTACCCGCGGAGCATCAGCGGCT 953  
5884 AACACCTTCCTTCAAGCGCGGCAATCAGCCGCTTACCCGCGGAGCATCAGCGGCT 5943  
954 TTGTAAGCGCATGAGCGCTGACGCGCGTGGCGGCGGCGGCGGAGGATTTGGGAACAAGAC 1013  
5944 TTGTAAGCGCATGAGCGCTGACGCGCGTGGCGGCGGCGGCGGAGGATTTGGGAACAAGAC 6003  
1014 CCGCTTCAAGCGCGCTTGGAGCCGCAACCGTTATGCGAATCTTTCGGGACCCCGATTGTC 1073  
6004 CCGCTTCAAGCGCGCTTGGAGCCGCAACCGTTATGCGAATCTTTCGGGACCCCGATTGTC 6063  
1074 GGGCTTCGCGCGTGTGAGTATCTACAAGAAGAGCGGACGCGACCGCAGCAAGAT 1133  
6064 GGGCTTCGCGCGTGTGAGTATCTACAAGAAGAGCGGACGCGACCGCAGCAAGAT 6123  
1134 TGAGGTTACCGCATTCAGCGCGGACCGCATCAGCTCCGCGCGGTGAGCTTATGCGG 1193  
6124 TGAGGTTACCGCATTCAGCGCGGACCGCATCAGCTCCGCGCGGTGAGCTTATGCGG 6183  
1194 ACCGATTCATCAGACCGCGCTGAGTGTATGAGCTTCAGCGCTGTGAGCGGACGAGGCGG 1253  
6184 ACCGATTCATCAGACCGCGCTGAGTGTATGAGCTTCAGCGCTGTGAGCGGACGAGGCGG 6243  
1254 CCGCAAGGCGCTTCCGCGGGGCAAGCCATTCTGCGCGCATGAGCAAGCTGTACTGGA 1313  
6244 CCGCAAGGCGCTTCCGCGGGGCAAGCCATTCTGCGCGCATGAGCAAGCTGTACTGGA 6303  
1314 GTGTGGCGCGCTCATGACTTGAAGCGCGGGAAGATCATGATCAAGGACTTTCACGCTG 1373  
6304 GTGTGGCGCGCTCATGACTTGAAGCGCGGGAAGATCATGATCAAGGACTTTCACGCTG 6363  
1374 CCGTCCGCGGGAAGTGTGTGACCCGCTCCGACCTTGGGCGACGACGAGGACGTCACAGT 1433  
6364 CCGTCCGCGGGAAGTGTGTGACCCGCTCCGACCTTGGGCGACGACGAGGACGTCACAGT 6423  
1434 CAGCATGGCGGCGCATTCGACGAAGTTCGTTGCGGAGCGATCTTCAACAAGATCAGGACGC 1493



QY 693 GCGGAATTGGCGGCTACGTCGGCGGGAGAGCCCTTACGGCTTGCAGCTTGTTCGGA 752  
 Db 735 GCGGAATTGGCGGCTACGTCGGCGGGAGAGCCCTTACGGCTTGCAGCTTGTTCGGA 794  
 QY 753 GACGAGAGATCAGCGGCAACGGCCGATGTCATGTCATCAACAAGTTAGGCA 812  
 Db 795 GACGAGAGATCAGCGGCAACGGCCGATGTCATGTCATCAACAAGTTAGGCA 854  
 QY 813 CTCGACACACTCCCTTACCGGACCCCTTGAAGTTCGAGCCGACGTAATCCGTTGTGTG 872  
 Db 855 CTCGACACACTCCCTTACCGGACCCCTTGAAGTTCGAGCCGACGTAATCCGTTGTGTG 914  
 QY 873 GCGGAGATCAAGAGCAACAACCTTCCCTTAAAGCCGGGAGTAAACCCGCAATCA 932  
 Db 915 GCGGAGATCAAGAGCAACAACCTTCCCTTAAAGCCGGGAGTAAACCCGCAATCA 974  
 QY 933 CCGGAGAGATCAAGAGCAACAACCTTCCCTTAAAGCCGGGAGTAAACCCGCAATCA 992  
 Db 975 CCGGAGAGATCAAGAGCAACAACCTTCCCTTAAAGCCGGGAGTAAACCCGCAATCA 1034  
 QY 993 CGAGACATTTGGGAGAGAGAGCCCTTCAAGCCCTGGGACCCGGCAACCGTTATGCAAT 1052  
 Db 1035 CGAGACATTTGGGAGAGAGAGCCCTTCAAGCCCTGGGACCCGGCAACCGTTATGCAAT 1094  
 QY 1053 CTTTCGGGACCCCGGCTATTGCGGGCTTGGCGGCTGAGGTATCTACAGAGAGAGCCGGA 1112  
 Db 1095 CTTTCGGGACCCCGGCTATTGCGGGCTTGGCGGCTGAGGTATCTACAGAGAGAGCCGGA 1154  
 QY 1113 CCGGACCCCGGCAACGAGAGATTGAGGGTTACCGGATTCAGCGGACCCGATTCAGCGCTCG 1172  
 Db 1155 CCGGACCCCGGCAACGAGAGATTGAGGGTTACCGGATTCAGCGGACCCGATTCAGCGCTCG 1214  
 QY 1173 GCGGATGAGCTGATTTGGCGGACCGATCATCGAGCCCGCTGATGATATAGCTTCAGCG 1232  
 Db 1215 GCGGATGAGCTGATTTGGCGGACCGATCATCGAGCCCGCTGATGATATAGCTTCAGCG 1274  
 QY 1233 GTGGTTGGAGCGAGGGGGCGGCAAGGGGCTTCCCGGGGCAAGCCATTCTGTCGCG 1292  
 Db 1275 GTGGTTGGAGCGAGGGGGCGGCAAGGGGCTTCCCGGGGCAAGCCATTCTGTCGCG 1334  
 QY 1293 CATGAGACACTGCTACTGCGAGTGTGGCGCGCTGATGACTTCGAGAGCGGCGGAGAGATC 1352  
 Db 1335 CATGAGACACTGCTACTGCGAGTGTGGCGCGCTGATGACTTCGAGAGCGGCGGAGAGATC 1394  
 QY 1353 GATCAAGAGCTTCTACCGCTGCGGCTGCGGAGGTGTCGAGACCCCTGCGCACTGGGCA 1412  
 Db 1395 GATCAAGAGCTTCTACCGCTGCGGCTGCGGAGGTGTCGAGACCCCTGCGCACTGGGCA 1454  
 QY 1413 GCAGGAGGACGCTGCAACGTCAGATGCGGCACTGCAACAAGTTGTCGGAAGCAT 1472  
 Db 1455 GCAGGAGGACGCTGCAACGTCAGATGCGGCACTGCAACAAGTTGTCGGAAGCAT 1514  
 QY 1473 CTTCAACAAGATAGGAGCGGCGCAAGGCGCAAGAGAGCTTGGCGCTTGTGGAGAGC 1532  
 Db 1515 CTTCAACAAGATAGGAGCGGCGCAAGGCGCAAGAGAGCTTGGCGCTTGTGGAGAGC 1574  
 QY 1533 CGCCGACGCTTGGGAGGCTACTAGGGGCTGAGAGAGGGGAGAGCGGCGCAACT 1592  
 Db 1575 CGCCGACGCTTGGGAGGCTACTAGGGGCTGAGAGAGGGGAGAGCGGCGCAACT 1634  
 QY 1593 TGTGGGAGGCGGCGGCGGCGCTTGAAGAGCTGTAGAGAGCGGCGGCG 1652  
 Db 1635 TGTGGGAGGCGGCGGCGGCGCTTGAAGAGCTGTAGAGAGCGGCGGCG 1694  
 QY 1653 AGGAGCTTACGAGGAGCCGCTTGGAGAGCACTTCCGAGAGCAACAGGAGCGCTGAC 1712  
 Db 1695 AGGAGCTTACGAGGAGCCGCTTGGAGAGCACTTCCGAGAGCAACAGGAGCGCTGAC 1754  
 QY 1713 GGTCCGGGAGAGGGGCGGAGAGGGGCTTGGCGAATTTGAAGCCGCGGAAGCCCGGA 1772  
 Db 1755 GGTCCGGGAGAGGGGCGGAGAGGGGCTTGGCGAATTTGAAGCCGCGGAAGCCCGGA 1814  
 QY 1773 GTTGCCTTGAACAATGTTCCCGAGAGAGCGGAGAGCGGCGCTTAAGTC 1832

Db 1815 GCTTCCCTTACCAATGTTCCCGGAGAGCGGAGAGCTGAGCCGCGGCTTAAGTC 1874  
 QY 1833 GTGGTGGGGGCGGCGCTACATAGACGACAGCCGCTTGTGTGGGCTCTTCGATACAA 1892  
 Db 1875 GTGGTGGGGGCGGCGCTACATAGACGACAGCCGCTTGTGTGGGCTCTTCGATACAA 1934  
 QY 1893 GATCGTTTTCACGAGTACGCTACGGGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1952  
 Db 1935 GATCGTTTTCACGAGTACGCTACGGGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1994  
 QY 1953 TTGATACCTGTTGGGAGAGCGGCGGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2012  
 Db 1995 TTGATACCTGTTGGGAGAGCGGCGGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2054  
 QY 2013 GGAAGAGGTAGCGGCTAG 2031  
 Db 2055 GGAAGAGGTAGCGGCTAG 2073  
 RESULT 5  
 AX114861  
 LOCUS AX114861 5711 bp DNA linear PAT 11-MAY-2001  
 DEFINITION Sequence 21 from Patent WO0129208.  
 ACCESSION AX114861  
 VERSION AX114861.1 GI:14031803  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 synthetic construct.  
 artificial construct.  
 REFERENCES  
 1 (bases 1 to 5711).  
 Kuenh, R., von Melchner, H. and Altschmid, J.  
 Conditional gene trapping construct for the disruption of genes  
 Patent: WO 0129208-A 21 26-Apr-2001;  
 JOURNAL Artemis Pharmaceuticals GmbH (DE); Frankgen Biotechnologie AG (DE)  
 FEATURES  
 location/Qualifiers  
 source 1..5711  
 BASE COUNT 1329 a 1511 c 1617 g 1254 t  
 ORIGIN  
 Query Match 88.5%; Score 1796.6; DB 6; Length 5711;  
 Best Local Similarity 99.2%; Pred. No. 0;  
 Matches 1805; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
 QY 213 GATGTGAGGTAGCGGCGGCTTACGACCGCTGAGTGGCGGAGCGGAGAGATAGCAGTGC 272  
 Db 1001 GGTGAGACGCTACGCGGCTTACGACCGTCACTGCGGAGCGGAGAGATTCGAGCGC 1060  
 QY 273 AGCAAGCCAGGAGACACAGCGTACGCGCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 332  
 Db 1061 AGCAAGCCAGGAGACACAGCGTACGCGCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1120  
 QY 333 AGTCCAGGCGGAGCGGCGGCTTACGCTGCGGCAATTTCAAGAGAGAGAGAGAGAGAGAG 392  
 Db 1121 AGTCCAGGCGGAGCGGCGGCTTACGCTGCGGCAATTTCAAGAGAGAGAGAGAGAGAGAG 1180  
 QY 393 GTGCGGTTCCGGGAGCGGAGCGGCGGAGTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 452  
 Db 1181 GTGCGGTTCCGGGAGCGGAGCGGCGGAGTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1240  
 QY 453 CGGCGGCTCAACATGATCATTTGTATGAGCTGTGCGGCTTCTGCGGCTTAAGAGTCAAT 512  
 Db 1241 CGGCGGCTCAACATGATCATTTGTATGAGCTGTGCGGCTTCTGCGGCTTAAGAGTCAAT 1300  
 QY 513 GAGCGGATTTCCGATTTGCTCGGAATTCCTGCGGCTTGGGCGTGAAGATTTGTTCACTCA 572  
 Db 1301 GAGCGGATTTCCGATTTGCTCGGAATTCCTGCGGCTTGGGCGTGAAGATTTGTTCACTCA 1360  
 QY 573 GGAAGGCTTTCGGGAGGAG 632

Dh 1361 GGAAGCGCTCTCCGGCAGGAGAAAGCTATGAGCTGATCTGATATATGCGGCTGCA 1420  
Qy 633 CGCGTCGACAAAGATTTCTGCTAAAGTCCGCCGAAGATTTCTGACAGAGAACCTTCA 692  
Db 1421 CGCGTCGACAAAGATTTCTGCTAAAGTCCGCCGAAGATTTCTGACAGAGAACCTTCA 1480  
Qy 693 GCGCAATTTGGGCGGCTAGTCCGGGGAAGGCGCTTACGGCTTCGAGCTTGTTCGA 752  
Db 1481 GCGCAATTTGGGCGGCTAGTCCGGGGAAGGCGCTTACGGCTTCGAGCTTGTTCGA 1540  
Qy 753 GACGAAGAGATCACGCGCAACGGCCGAATGGTCAATGCTCATCAACAGTTAGCGCA 812  
Db 1541 GACGAAGAGATCACGCGCAACGGCCGAATGGTCAATGCTCATCAACAGTTAGCGCA 1600  
Qy 813 CTCGACCACTCCCTTACCAGACCTTGGAGTTGAGCCGCAAGTATCCGGTGTGTC 872  
Db 1601 CTCGACCACTCCCTTACCAGACCTTGGAGTTGAGCCGCAAGTATCCGGTGTGTC 1660  
Qy 873 GCGTGAATCAAGAGCAGCAAAACCTTCCCTTCAAGCGGGGAGTCAAGCGGCAATTC 932  
Db 1661 GCGTGAATCAAGAGCAGCAAAACCTTCCCTTCAAGCGGGGAGTCAAGCGGCAATTC 1720  
Qy 933 CCGCGGAGCATCACGCGGCTTTGTAAGCGCATGAGACGCTGACGCGCTGCGACCGGAG 992  
Db 1721 CCGCGGAGCATCACGCGGCTTTGTAAGCGCATGAGACGCTGACGCGCTGCGACCGGAG 1780  
Qy 993 CGAGACGATTTGGGAAGAACCGCTTCAAGCGCTTGGGACCGCGGCAACCGTTATCGCA 1052  
Db 1781 CGAGACGATTTGGGAAGAACCGCTTCAAGCGCTTGGGACCGCGGCAACCGTTATCGCA 1840  
Qy 1053 CTTTGGGAGCCGCGTATTTGGGCGCTTGGCGCTGAGTGAATCTCAAGAAAGAACCGGA 1112  
Db 1841 CTTTGGGAGCCGCGTATTTGGGCGCTTGGCGCTGAGTGAATCTCAAGAAAGAACCGGA 1900  
Qy 1113 CGGACGCGGACACGAAAGATTGAGGTTACCGCATTCAGCGGACCGCATCGCTCG 1172  
Db 1901 CGGACGCGGACACGAAAGATTGAGGTTACCGCATTCAGCGGACCGCATCGCTCG 1960  
Qy 1173 GCGGTCGACGTTGATTTGGGACCGCATTCAGCGGACCGCATCGCTCG 1232  
Db 1961 GCGGTCGACGTTGATTTGGGACCGCATTCAGCGGACCGCATCGCTCG 2020  
Qy 1233 GTGTTGGAGCGGAGGCGGCGGCAAGGAGGCTTCCGCGGAGCAAGCATCTGCGGC 1292  
Db 2021 GTGTTGGAGCGGAGGCGGCGGCAAGGAGGCTTCCGCGGAGCAAGCATCTGCGGC 2080  
Qy 1293 CATGAGACAGCTGTACTGCGAGTGTGGCGCGCTCATGATTCGAAAGCGGCGGAAGATC 1352  
Db 2081 CATGAGACAGCTGTACTGCGAGTGTGGCGCGCTCATGATTCGAAAGCGGCGGAAGATC 2140  
Qy 1353 GATCAAGACGCTTACCGCTGCGGTCGCGGAAGTGTGTGCAACCGCTCGGACCTGGGCA 1412  
Db 2141 GATCAAGACGCTTACCGCTGCGGTCGCGGAAGTGTGTGCAACCGCTCGGACCTGGGCA 2200  
Qy 1413 GCACGAAGGACGTCAGAGTGTGGGCGGCTCATGCAAGTTCGTTGCGGAAGCAT 1472  
Db 2201 GCACGAAGGACGTCAGAGTGTGGGCGGCTCATGCAAGTTCGTTGCGGAAGCAT 2260  
Qy 1473 CTTCAACAGATCAGGACGCGGCAAGGCGGCAAGAGAGAGAGAGAGAGAGAGAGAG 1532  
Db 2261 CTTCAACAGATCAGGACGCGGCAAGGCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 2320  
Qy 1533 CGCGCGAGGCTTGGGCAAGGCTACAGAGGCGGCTGAGAAAGAGGCGGCAAGAGAGAG 1592  
Db 2321 CGCGCGAGGCTTGGGCAAGGCTACAGAGGCGGCTGAGAAAGAGGCGGCAAGAGAGAG 2380  
Qy 1593 TGTTCGAGGCGGCGGAGCGGCTTGAAGAGGCTGAGAGAGAGAGAGAGAGAGAGAGAG 1652  
Db 2381 TGTTCGAGGCGGCGGAGCGGCTTGAAGAGGCTGAGAGAGAGAGAGAGAGAGAGAGAG 2440  
Qy 1653 AGGACCTTACGAGGAGAGCGGTTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1712  
Db 2441 AGGACCTTACGAGGAGAGCGGTTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2500

Qy 1713 GCTCCGGACAGAGGCGGCGGAGAGAGCGGCTTGGCCGAACTTGAAGCGGCGGAGCGGCGGA 1772  
Db 2501 GCTCCGGACAGAGGCGGCGGAGAGAGCGGCTTGGCCGAACTTGAAGCGGCGGAGCGGCGGA 2560  
Qy 1773 GTTGGCCCTTGAACATGTTTCCCGAAGAGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAG 1832  
Db 2561 GCTTCCCTTGAACATGTTTCCCGAAGAGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAG 2620  
Qy 1833 GTTGGGCGGCGGCTGAGTACAGAGAGAGCGGCTTGTGTCGAGGCTTGTGAGACAA 1892  
Db 2621 GTTGGGCGGCGGCTGAGTACAGAGAGAGCGGCTTGTGTCGAGGCTTGTGAGACAA 2680  
Qy 1893 GATCGTTGACAGAGTGTACATGAGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1952  
Db 2681 GATCGTTGACAGAGTGTACATGAGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 2740  
Qy 1953 TTGATACAGTGGGCGGAGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2012  
Db 2741 TTGATACAGTGGGCGGAGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2800  
Qy 2013 GGAAGAGTGAAGCGGCTAG 2031  
Db 2801 GGAAGAGTGAAGCGGCTAG 2819  
  
RESULT 6  
PCL14670  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source  
1. 5715  
/organism="Cloning vector pSET152"  
/specific host="Streptomyces sp."  
/db\_xref="taxon:13260"  
/lab\_host="Escherichia coli"  
/gene="lacZa"  
/complement(163..507)  
/gene="lacZa"  
/complement(163..507)  
/function="blue-white screening"  
/codon\_start=1  
/transl\_table=11  
/product="lacZ alpha peptide"  
/protein\_id="CAC93946.1"  
/db\_xref="gi:17974210"  
/translation="MITSISRAADPLESTCPSLALAAVYLORRDWENPGVTOINRL  
AAHPPASMRNSSEARTRPQQQLBSLNGEWRIMRYFLTLTLCGISHRINPMSSTSG  
IGSADAKCRST"  
complement(508..509)  
/note="pUC18"  
/replace="ggt"  
  
variation







```

gene
CDS
    /translation="MTKTRLERFVAVGLEERASDDGRISMRYAYRNFNLSHDLGCF
    RERIVGAGAPSLRQNDVYATFNHHDRLSLGRSCTLRVGEDREGWMEIDLPDTTV
    GRDVAKLKRGDSSFTFRVLDGGRRADDDPETGLPVEITAMDEVGPPVNP
    APTTQASLRSIETLRIIGFAPPTERDSQPGDVAAPASIHFDARSLVALSK"
    4893..6071
    /gene="36"
    /protein_id="CAAO7110.1"
    /db_xref="GI:394745"
    /function="major capsid protein"
    /codon_start=1
    /transl_table=11
    /product="gp36"
    /protein_id="CAAO7106.1"
    /db_xref="GI:394745"
    /db_xref="SPTREMBL:O92XW6"
    /translation="MDATILSANFEAREERATAELRLTDEFAKEMTAREKEERLT
    TAVADFGRIKQIDAIKADAVTSLSLGSGSGAORSADHDDDAVLRANLIGAR
    SFEFAERKDRGKAGNPNVSLPTLNGQLAQAVERAIRMGCATTTSDANPMDET
    IGRATAGIVGETAEIPESEYPATTOBSKGFRTIGFASVSYEFADQVLDLVGFLVSD
    AGPAIDGAKGRHELTGTGTGQPGILTDVYGANAFAEGADADSKVSADALIDLFEHPS
    AYRKNKFEVNDLRAQMRKRLDANQOYLMOGALTVGAPDTENGKVEVETDDGMPADKY
    LEADLSKYRVRERFAGSLRVDRSVDKFTSTQIYVRFQRADGLLVDAKAKVLTIVPA
    "
    5226..6068
    /gene="36"
    /product="gp36"
    /db_xref="GI:394745"
    /gene="37"
    6144..6512
    /gene="37"
    /codon_start=1
    /transl_table=11
    /product="gp37"
    /protein_id="CAAO7107.1"
    /db_xref="GI:394745"
    /db_xref="SPTREMBL:O92XB0"
    /translation="MAYATIELRALDGLDLSALPSDELLSDAIDFSVETVAYCGRK
    MDTADDPETPTIRMCVRIARQYLDHVSRIEDRALQLOSERGSIQLAOGANMPT
    LPEVNANKLTVRVRPEITFM"
    6519..6956
    /gene="38"
    6519..6956
    /gene="38"
    /codon_start=1
    /transl_table=11
    /product="gp38"
    /protein_id="CAAO7108.1"
    /db_xref="GI:394745"
    /db_xref="SPTREMBL:O92XA9"
    /translation="MALIFNAKVRLEFALKANVPDQVCTFAETGDSNRKROWLGAT
    VDDDLAPVAMRSKAKPTNVTGYVHAHVTTGNGNPVDAEPRAVYGIIDVYKAACAVNA
    DLVSYPGIMDVRPESASVESTETTDGATSLTVRVRGRVYQ"
    6989..7975
    /gene="39"
    6989..7975
    /gene="39"
    /function="major tail protein"
    /codon_start=1
    /transl_table=11
    /product="gp39"
    /protein_id="CAAO7109.1"
    /db_xref="GI:394745"
    /db_xref="SPTREMBL:O92XA8"
    /translation="MALDASIGIGREDIYGLSGVESEGOADSMKOTVPISSVCF
    RAGMOTARADRRNVMNGEGELEIDVLDAGAGSLTNAFRTVTTDGVGKTVLET
    SDVSEAFETIAQVNPQVDSKVAFKHLGCVATWMSLTAEEVAKLAVTFDFDATH
    TSDPAIEPTYPAAVYDWTRTAVELSRAGSVYEPDATISLELTGELGMKIDRRFN
    GSELKKKVRNALPTVEGLEEFSAASLGLYDAFVSGEVCSFKATFPGILGSLV
    EAPAIQFTGSEPAATDEVITLHNPFRVLDPGDGVTAIKILTYEPGTPVP"
    7975..8406
    /gene="40"
    7975..8406
    /gene="40"
    /gene="40"

```

```

Query Match      88.5%; Score 1796.6; DB 7; Length 41489;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 1805; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

gene
    /codon_start=1
    /transl_table=11
    /product="gp40"
    /protein_id="CAAO7110.1"
    /db_xref="GI:394745"
    /db_xref="SPTREMBL:O92XA7"
    /translation="MAORSAVYIOYDGFQFORNVBALRDKELNKAVERANKSGEVL
    IPOAKHESPDGHRDPSKSRVRPKLDSIKYTKASAKGAVITAGSARPYAAIHFG
    YRKNIISANREPIYRMARSDVVAATYERRIAAVYEKYLE"
    8406..8878

```



```

Db 39284 CTTTCGGGACCCGCTATTGCGGGCTTCGCCGTGAGTGATCTACAAAGAAAGACCCGA 39343
Oy 1113 CGGCACGGCGACACAGAGATTGAGGGTTAACCGCATTTACGCCGACCCGATACCGTCCG 1172
Db 39344 CGGCACGGCGACACAGAGATTGAGGGTTAACCGCATTTACGCCGACCCGATACCGTCCG 39403
Oy 1173 GCGGCTGACGCTTGAATTCGCGACCGATCATGAGCCCGCTGAGTATGACCTTCAGC 1232
Db 39404 GCGGCTGACGCTTGAATTCGCGACCGATCATGAGCCCGCTGAGTATGAGCTTCAGC 39463
Oy 1233 GTGTTGAGACGGCAGGGGGCGCGCAAGGGGCTTTCCGGGGGCAAGCCATTCTGTCGC 1292
Db 39464 GTGTTGAGACGGCAGGGGGCGCGCAAGGGGCTTTCCGGGGGCAAGCCATTCTGTCGC 39523
Oy 1293 CATGACAAAGCTTACTCGAGTGTGGGGCCCTCATGACTTGGAAAGCGGGGGAAGATC 1352
Db 39524 CATGACAAAGCTTACTCGAGTGTGGGGCCCTCATGACTTGGAAAGCGGGGGAAGATC 39583
Oy 1353 GATCAAGGACTCTTACCCGCTGCGCGGAGAGGTGCTGACCCGCTCCGACCTGGGCA 1412
Db 39584 GATCAAGGACTCTTACCCGCTGCGCGGAGAGGTGCTGACCCGCTCCGACCTGGGCA 39643
Oy 1413 GCACGAAGGACGTCAACAGTACATGGCGGCACTCGACAAGTTCTGTCGGAAAGCAT 1472
Db 39644 GCACGAAGGACGTCAACAGTACATGGCGGCACTCGACAAGTTCTGTCGGAAAGCAT 39703
Oy 1473 CTTCAACAAGATCAGGCAAGCGCGGAAGGAGCAAGAGAGCTGGGGCTTCTGGGAAGC 1532
Db 39704 CTTCAACAAGATCAGGCAAGCGCGGAAGGAGCAAGAGAGCTGGGGCTTCTGGGAAGC 39763
Oy 1533 CGCCCGACGCTTTCGGCAAGCTCTGAGGCGCCCTGAGAAAGAGCGGGAACGGCAACT 1592
Db 39764 CGCCCGACGCTTTCGGCAAGCTCTGAGGCGCCCTGAGAAAGAGCGGGAACGGCAACT 39823
Oy 1593 TGTTCGGAGCGCGCGGAGCGCCCTTGAAGAGCTGTACGAAGAACCGCGCGGC 1652
Db 39824 TGTTCGGAGCGCGCGGAGCGCCCTTGAAGAGCTGTACGAAGAACCGCGCGGC 39883
Oy 1653 AGGAGCTTACGAGGAGCGCGTGGGAGAGAGACTTCCGGAAGCAACAGCAGCGCTGAC 1712
Db 39884 AGGAGCTTACGAGGAGCGCGTGGGAGAGAGACTTCCGGAAGCAACAGCAGCGCTGAC 39943
Oy 1713 GCTCCGAGAGCAAGGGGGCGGAAGAGCGGCTTCCGAAGTTGAAGCGCGGAGCCCGAA 1772
Db 39944 GCTCCGAGAGCAAGGGGGCGGAAGAGCGGCTTCCGAAGTTGAAGCGCGGAGCCCGAA 40003
Oy 1773 GTTCCGCTTGAACAATGGTTCCTCCGAAGACGCGGACGCTGACCGCGCCCTTAAGTC 1832
Db 40004 GTTCCGCTTGAACAATGGTTCCTCCGAAGACGCGGACGCTGACCGCGCCCTTAAGTC 40063
Oy 1833 GTGGTGGGGGGCGGGGTAGTAGAGCAAGGGCGTGTCTGCGGGCTTCTGTAGACAA 1892
Db 40064 GTGGTGGGGGGCGGGGTAGTAGAGCAAGGGCGTGTCTGCGGGCTTCTGTAGACAA 40123
Oy 1893 GATCGTTTTCACGAGTACGACTACGAGGCGAGGGGAGGAGCCCATCGAAGAGCGGC 1952
Db 40124 GATCGTTTTCACGAGTACGACTACGAGGCGAGGGGAGGAGCCCATCGAAGAGCGGC 40183
Oy 1953 TTTCGATCAAGTGTGGGCAAGCCCGCAGACGACGACGAAGAAGAGCGCCAGAGCGGAC 2012
Db 40184 TTTCGATCAAGTGTGGGCAAGCCCGCAGACGACGACGACGAAGAAGAGCGCCAGAGCGGAC 40243
Oy 2013 GGAACAGCTAGCGGCGTAG 2031
Db 40244 GGAACAGCTAGCGGCGTAG 40262

```

```

RESULT 8
LOCUS EVU84006 5967 bp DNA circular SYN 03-MAY-1997
DEFINITION Expression vector pBSII-LUCINT firefly luciferase (LUCINT),
beta-galactosidase (lacZ) and beta-lactamase (ampr) genes, complete
cds and lac operon promoter sequence.

```

```

ACCESSION U84006
VERSION U84006.1 GI:2071944
KEYWORDS
SOURCE Expression vector pBSII-LUCINT.
ORGANISM Expression vector pBSII-LUCINT.
REFERENCE Artificial sequences; vectors.
AUTHORS 1 (bases 1 to 5967)
TITLE Mankin,S.L., Allen,G.C. and Thompson,W.F.
INTRODUCTION Introduction of a plant intron into the Luciferase Gene of Photinus
pyralis
JOURNAL Plant Mol. Biol. Rep. (1997) In press
REFERENCE 2 (bases 1 to 5967)
AUTHORS Mankin,S.L., Allen,G.C. and Thompson,W.F.
TITLE Direct Submission
JOURNAL Submitted (06-JAN-1997) Botany, NC State Univ., Box 7612, Raleigh,
NC 27695-7612, USA
FEATURES
source
1..5967
/organism="Expression vector pBSII-LUCINT"
/db_xref="taxon:56405"
/plasmid="pluk07"
/note="chimeric plasmid consisting of a pbs-type plasmid
(pluk07, dsDNA) derived from the LUC* gene (a synthetic
cDNA gene) by the addition of a plant intron sequence"
52..851
/note="CamV 35S promoter"
875..3052
/gene="LUCINT"
join(875..1029,1219..2716)
/gene="LUCINT"
/note="luciferin 4-monooxygenase: LUC* with PIV2 intron"
/codon_start=1
/product="firefly luciferase"
/protein_id="AAB53627.1"
/db_xref="GI:2071945"
/translation="MEDAKNIKRGAPPEYPLEDGTAGEBOLHKAMKRYALVPTIAFTD
AHIEVNTYAEFEFMSVRLAEKRYGKNTNRIYVCSNSLOFEMPVYALFIGNV
APANDYNREILNMSNISQPIYVYVSKKGLTKLNNOKKLEIITOKIILMSKTYOG
FOSMTFVTSHPGPFNEIDFVPSFDRKDTALIMNSGSGTGLKRGALPRTACVR
FSHADPIFGNOIIPDTALISVPHHGGMTTGLYLCGRVYLVAMRFEELFLRS
LDQYIQSALVPLTFSFAKSTLIDKXDLNMLHIASGAPLSKEVEBAVRFLP
GRQGYGLTETTSALITPBGDKGAVGVYVFEAKVVDLDTGKTIGVNOGELCV
RGPMTSGVYNNPEATNALIDKDGMLHSGDIAVWDEHEFIVDLKSLIKKGYOVA
PALESTILQHNPIFDAGYAGLPDDAGELPRAVYVLEBKMTKEIYDIYASQVTT
AKLKGAVFVDEVPKGLDKARKIRILIKKGGKSL"
1030..1218
/gene="LUCINT"
/note="PIV2 chimeric intron"
2786..3052
/gene="LUCINT"
/standard_name="nopoline synthase 3' region"
/note="nos"
2816..2821
/gene="LUCINT"
3146..3355
/gene="lacZ"
3146..3355
/gene="lacZ"
3146..3355
/codon_start=1
/transl_table=11
/product="beta-galactosidase"
/protein_id="AAB53629.1"
/db_xref="GI:2071947"
/translation="NAVVLORDMPCVGTQILRLAHPFPAWNSSEARTDRPSQO
LRSLNGEWEIYNNVILKFLALNFC"
3895..4755
/gene="ampr"
3895..4755
/gene="ampr"
/function="ampicillin resistance marker"
/codon_start=1
/transl_table=11

```

	/product="beta-lactamase"	
	/protein_id="AAB53628.1"	
	/db_xref="GI:2071946"	
	/translation="MSIQHFRVALIPFPAFLPYFAHPETLVKVKDAEDLGARVGY	
	LELDLNSGKILIESRPEREEPRMSTFKVLLCGAVLSRIDAOEOLGRIRIHYSONDLVIV	
	YSPVEKHLLTDMTVIRELCSSAITMSDNTANILLTLTGSRKEILTAFLHNMGDHTRID	
	DMEPEELNGLIFPNDERTDTMPYAAMATTLRKLTGLTLLASROOLIIPWMEADKVAGPI	
	LRSALPLKHWFIADKSGAGERGSRGIIIALGPDKRSRIIVLYITTSQATMDERNQIIR	
	EGASLIKHM"	
	5790..5912	
promoter	/note="lac operon promoter"	
BASE COUNT	1686 a 1349 c 1416 g 1516 t	
ORIGIN		
Query Match	9.9%; Score 200.4; DB 12; Length 5967;	
Best Local Similarity	97.1%; Pred. No. 1.3e-27;	
Matches 204; Conservative	0; Mismatches 6; Indels 0; Gaps 0;	
Oy 30	TACGTAACTTTTCGCTCTTACCTTTGGATATATATATAATTATTCATTAACTTAGTAGTA 89	
Db 1027	TACGTAACTTTTCGCTCTTACCTTTGGATATATATAATTATTCATTAACTTAGTAGTA 1086	
Oy 90	AATATAATATTTCAAAATATTTTTTTCAAAATAAGAAGCTAGTATATAGCAATTGCCTTTT 149	
Db 1087	ATATATATATTTCAAATATTTTTTTTTCAAAATAAGAAGTATATATAGCAATTGCCTTTT 1146	
Oy 150	CCTGACTTTTATAAGTGTGATATATTTTAAATTTTAACTTTTCTAATATATGACCAGAAATTT 209	
Db 1147	CTGTAGTTTATAAGTGTGATATATTTTAAATTTTAACTTTTCTAATATATGACCAGAAATTT 1206	
Oy 210	GTTGATGTGCAGGTACGCCGGCGCTACGA 239	
Db 1207	GTTGATGTGCAGGTACGCCGAATFACTGCA 1236	
RESULT 9		
LOCUS	AX259239 912 bp DNA linear PAT 26-OCT-2001	
DEFINITION	Sequence 37 from Patent WO0173087.	
ACCESSION	AX259239	
VERSION	AX259239.1 GI:16508485	
KEYWORDS		
SOURCE		
ORGANISM	synthetic construct.	
REFERENCE	synthetic construct	
AUTHORS	artificial sequences.	
JOURNAL	Hohn,T., Stavolone,L., de Haan,P.T., Ligon,H.T. and Kononova,M. Cestrum yellow leaf curling virus promoters Patent: WO 0173087-A 37 July 04-OCt-2001; Syngenta Participations AG (CH) Location/Qualifiers	
FEATURES	1..912	
source	1..912	
BASE COUNT	245 a 245 c 240 g 182 t	
ORIGIN		
Query Match	9.8%; Score 198.6; DB 6; Length 912;	
Best Local Similarity	95.8%; Pred. No. 2.1e-27;	
Matches 204; Conservative	0; Mismatches 9; Indels 0; Gaps 0;	
Oy 21	CGGGGTGATACGTAACTTTCTGCTTACCTTTGATATATATATAATTAATTCATTAA 80	
Db 265	CCGGGGCGCTACGTAACTTTCTGCTTACCTTTGATATATATAATTAATTCATTAA 324	
Oy 81	TTAGTAGTAATATATATTTTCAAATATTTTTTCAAAAAATAAGAATGTAGTATAGCA 140	
Db 325	TTAGTAGTAATATATATTTTCAAATATTTTTTCAAAAAATAAGAATGTAGTATAGCA 384	
Oy 141	ATTGCTTTTCTGTATTAAGTGTGATATATTTTAACTTTTCTAATATATAGCA 200	
Db 385	ATTGCTTTTCTGTATTAAGTGTGATATTTTAACTTTTCTAATATATAGCA 444	

OY	201	CCAAATTGTTGTATGCAGTACGGGGTCG	233
Db	445	CCAAATTGTTGTATGCAGTGCGAGGCGC	477
RESULT 10			
LOCUS	AX259244	1577 bp	DNA
DEFINITION	Sequence 42 from Patent WO0173087.		Linear
ACCESSION	AX259244		PAT 26-OCT-2001
VERSION	AX259244.1		
KEYWORDS	GI:16508490		
SOURCE			
ORGANISM	synthetic construct. synthetic construct artificial sequences.		
REFERENCE	1		
AUTHORS	Hohn,T., Stavolone,L., de Haan,P.T., Ligon,H.T. and Kononova,M.		
TITLE	Cestrum yellow leaf curling virus promoters		
JOURNAL	Patent: WO 0173087-A 42 04-Oct-2001; Syngenta Participations AG (CH)		
FEATURES	Location/Qualifiers		
source	1..1577		
	/organism="synthetic Construct" /db_xref="taxon:32630"		
	/note="artificial sequence Cmpc-syngFP1-nos expression cassette"		
BASE COUNT	450 a 372 c 390 g 365 t		
ORIGIN			
Query Match	9.8%; Score 198.6; DB 6; Length 1577;		
Best Local Similarity	95.8%; Pred. No.2,3e-27;		
Matches	204; Conservative 0; Mismatches 9; Indels 0; Gaps 0.		
OY	21	CGGGTGCACTACGTAAGTTTCGCTTCACTTGGATATAATATAATATCATTA	80
Db	645	CGGGGGGGCGTAGTGAATTCGCTTCACTTGGATATAATATAATATCATTA	704
OY	81	TAGTAGTAATATAATATTTCAAATATTTTTTCCAAAATAAAGAATGAGTATATGACA	140
Db	705	TAGTAGTAATATAATATTTCCAATATTTTTTCCAAAATAAAGAATGAGTATATGACA	764
OY	141	ATTGCTTTTCGTACTATTATAAGTGTATATTTAATTTATATACCTTTCTAATATATGCA	200
Db	765	ATTGCTTTTCGTACTATTATAAGTGTATATTTAATTTAATTTAATTTCTAATATATGCA	824
OY	201	CCAAATTGTTGTATGCAGTACGGGGTCG	233
Db	825	CCAAATTGTTGTATGCAGTGCGAGGCGC	857
RESULT 11			
LOCUS	AX259242	1618 bp	DNA
DEFINITION	Sequence 40 from Patent WO0173087.		Linear
ACCESSION	AX259242		PAT 26-OCT-2001
VERSION	AX259242.1		
KEYWORDS	GI:16508488		
SOURCE			
ORGANISM	synthetic construct. synthetic construct artificial sequences.		
REFERENCE	1		
AUTHORS	Hohn,T., Stavolone,L., de Haan,P.T., Ligon,H.T. and Kononova,M.		
TITLE	Cestrum yellow leaf curling virus promoters		
JOURNAL	Patent: WO 0173087-A 40 04-Oct-2001; Syngenta Participations AG (CH)		
FEATURES	Location/Qualifiers		
source	1..1618		
	/organism="synthetic Construct" /db_xref="taxon:32630"		
	/note="artificial sequence Cmps-syngFP1-nos expression cassette"		
BASE COUNT	468 a 383 c 400 g 367 t		



QY 150 CTGTAGTTTATAGTGTGATATTTTAACTTTCTAATATATGACCAAAATTT 209  
|||||  
DB 503 CTGTAGTTTATAGTGTGATATTTTAACTTTCTAATATATGACCAAAATTT 562  
|||||  
QY 210 GTTGATGTCAGGTA 224  
|||||  
DB 563 GTTGATGTCAGGTA 577  
|||||

## .RESULT 15

AR037156

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 195; Conservative

9.6%; Score 195; DB 6; Length 2633;

100.0%; Pred. No. 1.2e-26;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

0; Mismatches 0; Indels 0; Gaps 0;

Search completed: May 11, 2003, 21:26:57  
Job time : 6061.5 secs

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 14:41:33 ; Search time 511.5 seconds  
(without alignments)  
8941.952 Million cell updates/sec

Title: US-09-940-550A-10  
Perfect score: 2031  
Sequence: 1 atggcacaaggggtgtgac.....cggagacgtacgagcgtacg 2031

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: N\_Geneseq\_101002.\*  
2: /SID52/gcgdata/geneseq/geneseq-n-emb1/NA1980.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq-n-emb1/NA1981.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq-n-emb1/NA1982.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq-n-emb1/NA1983.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq-n-emb1/NA1984.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq-n-emb1/NA1985.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq-n-emb1/NA1986.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq-n-emb1/NA1987.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq-n-emb1/NA1988.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq-n-emb1/NA1989.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq-n-emb1/NA1990.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq-n-emb1/NA1991.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq-n-emb1/NA1992.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq-n-emb1/NA1993.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq-n-emb1/NA1994.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq-n-emb1/NA1995.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq-n-emb1/NA1996.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq-n-emb1/NA1997.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq-n-emb1/NA1998.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq-n-emb1/NA1999.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq-n-emb1/NA2000.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseq-n-emb1/NA2001.DAT.\*  
24: /SID52/gcgdata/geneseq/geneseq-n-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	2031	100.0	2031	24 ABK12556
2	2015	99.2	2031	24 ABK12555
3	2015	99.2	2031	24 ABK12557
4	1796.6	88.5	5711	22 AAD04935
5	1795	88.4	1842	22 AAH74879
6	1793.4	88.3	3401	11 AAO06904
7	1790.2	88.1	1842	22 AAH74880
8	1785.6	87.9	1839	22 AAH74881
9	1767	87.0	1833	22 AAH74882

10	198.6	9.8	912	22 AAD19829
11	198.6	9.8	1577	22 AAD19834
12	198.6	9.8	1618	22 AAD19832
13	198.6	9.8	2943	22 AAD19840
14	195	9.6	2001	22 AAD19830
15	195	9.6	2001	24 ABK15667
16	195	9.6	2725	22 AAD19835
17	195	9.6	2730	22 AAD19833
18	195	9.6	4072	22 AAD19841
19	195	9.6	4341	22 AAD19839
20	195	9.6	5534	17 AAT43137
21	195	9.6	12817	24 ABK15668
22	195	9.6	13274	24 ABK15666
23	194	9.6	1701	21 AAZ29120
24	194	9.6	3336	21 AAZ29121
25	194	9.6	3877	21 AAZ29123
26	192	9.5	1294	22 AAD19907
27	190.8	9.4	1691	22 AAD19910
28	190	9.4	1034	19 AAV54572
29	181.4	8.9	363	22 AAF32096
30	180.6	8.9	363	22 AAF32093
31	180.6	8.9	363	22 AAF32095
32	180.6	8.9	369	22 AAF32094
33	137.8	6.8	1512	22 AAD19908
34	137.8	6.8	2708	22 AAD19909
35	137.4	6.8	1851	22 AAD19912
36	137.4	6.8	2315	22 AAD19911
37	58.2	2.9	154902	24 ABQ88198
38	53.6	2.6	7752	24 ABQ88126
39	53.6	2.6	159400	24 ABQ88126
40	53	2.6	6641	24 ABL54335
41	53	2.6	6641	24 ABL32314
42	52.2	2.6	6118	24 ABN80110
43	51.8	2.6	1836	24 ABL55639
44	51.8	2.6	4654	22 AAS46780
45	51.8	2.6	4654	24 ABL34222

## ALIGNMENTS

RESULT 1  
ABK12556  
ID ABK12556 standard; DNA: 2031 BP.  
AC ABK12556;  
XX  
DT 18-JUN-2002 (first entry)  
XX  
DE DNA encoding novel phl C31 integrase, phlC31int\*INT.  
XX  
KW Phl C31 integrase; recombinase; transgenic; plant;  
KW agricultural food product; self-excising polynucleotide;  
KW gene; ds; phlC31int\*INT.  
XX  
OS Synthetic.  
XX  
PN WO200216609-A2.  
XX  
PD 28-FEB-2002.  
XX  
PF 27-AUG-2001; 2001WO-US26738.  
XX  
PR 25-AUG-2000; 2000US-227961P.  
XX  
PA (BADI ) BASF PLANT SCI GMBH.  
XX (MCKE/ ) MCKERSIE B.  
XX  
PI Mankin L;  
XX  
DR WPI: 2002-280939/32.  
XX  
PT New self-excising polynucleotides, useful for producing transgenic

Green fluorescent  
Promoter-reporter  
Promoter-reporter  
Ubp3(At)-synGFP1-N  
Beta-glucuronidase  
Glutathione-S-tran  
Promoter-reporter  
Promoter-reporter  
Ubp3(At)-GIC-NOS t  
Zmubi1-GIC-NOS term  
pMIGIT sequence 1  
Expression vector  
Binary vector PNOV  
Plasmid DV127 used  
Plasmid DV132 comp  
pCI vector based o  
Glucosamine-6-phos  
Insecticidal prote  
Insecticidal prote  
Insecticidal prote  
pTBN1 vector based  
pTBN6 vector based  
pCII vector based  
pC16 vector based  
Human osteoblast d  
Chemically treated  
Human osteoblast d  
Chemically treated  
Human immune syste  
Human chemically m  
AntiEpV second DNA p  
Tumour suppressor  
Human immune syste

filed  
instant  
US

PT plants, removing transgenes from these plants or crops (e.g. food  
PT commodities), and restricting the distribution of transgenes within the  
PT environment

XX Claim 5; Fig 7; 60pp; English.

CC The invention describes an isolated excisable polynucleotide, which  
CC comprises a desired trait polynucleotide and a recombinase polynucleotide  
CC operably linked to a promoter, all flanked by a pair of directly oriented  
CC recombination sites where the recombinase activity is regulatable. The  
CC self-excising polynucleotide is useful for producing transgenic plants,  
CC particularly for removing all transgenic nucleic acid sequences that  
CC encode a gene product from the transgenic plant, thus restoring the  
CC original genetic configuration of the genome. The polynucleotide is also  
CC useful in methods for eliminating unwanted nucleic acids in agricultural  
CC food products and for preventing the escape of certain transgenic traits  
CC into the environment. This sequence encodes the novel phic31 integrase  
CC phic31int\*1M, created for use in the self-excising polynucleotide  
CC described in the invention.

XX Sequence 2031 BP; 454 A; 558 C; 640 G; 379 T; 0 other;

Query Match 100.0%; Score 2031; DB 24; Length 2031;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Gaps 0;

Matches 2031; Conservative 0; Indels 0; Gaps 0;

QY 1 ATGCGACAAGGGGTTGTGACCGGGGTGATACGTTCTGCTTACCTTGATATA 60  
DB 1 ATGCGACAAGGGGTTGTGACCGGGGTGATACGTTCTGCTTACCTTGATATA 60  
QY 61 TAT 120  
DB 61 TAT 120  
QY 121 AAAGAATAGTAT 180  
DB 121 AAAGAATAGTAT 180  
QY 181 ATATATTTTAT 240  
DB 181 ATATATTTTAT 240  
QY 241 CGTACGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300  
DB 241 CGTACGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300  
QY 301 AACGAAGACAAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360  
DB 301 AACGAAGACAAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360  
QY 361 TTGCTGGGCGATTTCAAGCGAAGCGCGGCGGCGGCGGCGGCGGCGGCGG 420  
DB 361 TTGCTGGGCGATTTCAAGCGAAGCGCGGCGGCGGCGGCGGCGGCGGCGG 420  
QY 421 GAGTTGGAAGCGATCTTGAAGCAAGTCCGCGGCGGCGGCGGCGGCGGCGG 480  
DB 421 GAGTTGGAAGCGATCTTGAAGCAAGTCCGCGGCGGCGGCGGCGGCGGCGG 480  
QY 481 GAGCTTCCGCGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540  
DB 481 GAGCTTCCGCGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540  
QY 541 CTCGCGCTGGGCGTGAAGTGTTCCTCACTCAAGAAAGGCGGCGGCGGCGG 600  
DB 541 CTCGCGCTGGGCGTGAAGTGTTCCTCACTCAAGAAAGGCGGCGGCGGCGG 600  
QY 601 ATGAGACTGATTCACCTGATATGCGGCGTCAAGCGGCGGCGGCGGCGG 660  
DB 601 ATGAGACTGATTCACCTGATATGCGGCGTCAAGCGGCGGCGGCGGCGG 660  
QY 661 TCGGGGAAGATTTCTGACACGAAAGCTTCAGCGCGAATTCGCGGCGGCGG 720  
DB 661 TCGGGGAAGATTTCTGACACGAAAGCTTCAGCGCGAATTCGCGGCGGCGG 720

QY 721 AAGCGCCTTACGCGCTTCGAGCTTGTTCGGAGACGAAAGAGATCACGCCGCGCA 780  
DB 721 AAGCGCCTTACGCGCTTCGAGCTTGTTCGGAGACGAAAGAGATCACGCCGCGCA 780  
QY 781 ATGCTCATATGCTGATCAACAAGTTAGCGGCGGCGGCGGCGGCGGCGGCGG 840  
DB 781 ATGCTCATATGCTGATCAACAAGTTAGCGGCGGCGGCGGCGGCGGCGGCGG 840  
QY 841 GAGTTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900  
DB 841 GAGTTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900  
QY 901 CCGTTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960  
DB 901 CCGTTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960  
QY 961 CGCATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020  
DB 961 CGCATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020  
QY 1021 AAGCGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080  
DB 1021 AAGCGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080  
QY 1081 GCGGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140  
DB 1081 GCGGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140  
QY 1141 TACCGCATTCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200  
DB 1141 TACCGCATTCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200  
QY 1201 ATGAGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1260  
DB 1201 ATGAGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1260  
QY 1261 GGGCTTTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1320  
DB 1261 GGGCTTTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1320  
QY 1321 GCGGCTGATCTTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1380  
DB 1321 GCGGCTGATCTTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1380  
QY 1381 CGGAAGGTGTGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1440  
DB 1381 CGGAAGGTGTGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1440  
QY 1441 GCGGCTGATCTTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1500  
DB 1441 GCGGCTGATCTTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1500  
QY 1501 GAGGGAAGAGCGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1560  
DB 1501 GAGGGAAGAGCGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1560  
QY 1561 GCGGCTGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1620  
DB 1561 GCGGCTGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1620  
QY 1621 GCGGCTGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1680  
DB 1621 GCGGCTGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1680  
QY 1681 AAGGCTTCGGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1740  
DB 1681 AAGGCTTCGGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1740  
QY 1741 CTTGCGGAATTTGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1800  
DB 1741 CTTGCGGAATTTGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1800

QY 1801 GACGCGACGCTGACCCGACCGGCGCTTAAGTCGTGTGGGGGCGCGCTCAGTAGACGAC 1860  
DB 1801 GACGCGACGCTGACCCGACCGGCGCTTAAGTCGTGTGGGGGCGCGCTCAGTAGACGAC 1860  
QY 1861 AAGCGCGCTGCTGTCGGGCTCTTGTGACAGATCGTTGTCAGCACTACAGGCG 1920  
DB 1861 AAGCGCGCTGCTGTCGGGCTCTTGTGACAGATCGTTGTCAGCACTACAGGCG 1920  
QY 1921 AAGGCGCGACGCGCGCTTGTGACAGATCGTTGTCAGCACTACAGGCG 1980  
DB 1921 AAGGCGCGACGCGCGCTTGTGACAGATCGTTGTCAGCACTACAGGCG 1980  
QY 1981 GACGACGACGACGACGCGCGCTTGTGACAGATCGTTGTCAGCACTACAGGCG 2031  
DB 1981 GACGACGACGACGACGCGCGCTTGTGACAGATCGTTGTCAGCACTACAGGCG 2031  
RESULT 2  
ABK12555  
ID ABK12555 standard; DNA; 2031 BP.  
XX  
AC ABK12555:  
XX  
DT 18-JUN-2002 (first entry)  
XX  
DE DNA encoding novel phl C31 integrase, phlC31IntInt.  
XX  
KW Phl C31 integrase; recombinase; transgenic plant;  
KW agricultural food product; self-excising polynucleotide;  
KW gene; ds: phlC31IntInt.  
XX  
OS Synthetic.  
XX  
PN WO200216609-A2.  
XX  
PD 28-FEB-2002.  
XX  
PF 27-AUG-2001; 2001WO-US26738.  
XX  
PR 25-AUG-2000; 2000US-227961P.  
PA (BADI ) BASF PLANT SCI GMBH.  
PA (MCKE/) MCKERSIE B.  
PI Mankin L;  
XX  
DR WPI; 2002-280939/32.  
PT New self-excising polynucleotides, useful for producing transgenic  
PT plants, removing transgenes from these plants or crops (e.g. food  
PT commodities), and restricting the distribution of transgenes within the  
PT environment  
XX  
PS Claim 5; Fig 6; 60pp; English.  
XX  
CC The invention describes an isolated excisable polynucleotide, which  
CC comprises a desired trait polynucleotide and a recombinase polynucleotide  
CC operably linked to a promoter, all flanked by a pair of directly oriented  
CC recombination sites where the recombinase activity is regulatable. The  
CC self-excising polynucleotide is useful for producing transgenic plants,  
CC particularly for removing all transgenic nucleic acid sequences that  
CC encode a gene product from the transgenic plant, thus restoring the  
CC original genetic configuration of the genome. The polynucleotide is also  
CC useful in methods for eliminating unwanted nucleic acids in agricultural  
CC food products and for preventing the escape of certain phlC31 integrase  
CC into the environment. This sequence encodes the novel phlC31 integrase  
CC phlC31IntInt, created for use in the self-excising polynucleotide  
CC described in the invention.  
XX  
SQ Sequence 2031 BP; 451 A; 562 C; 640 G; 378 T; 0 other;

Matches 2021; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
QY 1 ATGCGACAAAGGGCTGTGACCGGGGTGATACGTAGTTTCGCTTACCTTGATATA 60  
DB 1 ATGCGACAAAGGGGTGTGACCGGGGTGATACGTAGTTTCGCTTACCTTGATATA 60  
QY 61 TATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 120  
DB 61 TATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 120  
QY 121 AAAGAAATAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 180  
DB 121 AAAGAAATAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 180  
QY 181 ATAACTTTTCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 240  
DB 181 ATAACTTTTCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 240  
QY 241 CGTCAGTCGCGGACGCGAGAAATAGCAGTCGACGACCCGACGACGACGACGAC 300  
DB 241 CGTCAGTCGCGGACGCGAGAAATAGCAGTCGACGACCCGACGACGACGACGAC 300  
QY 301 AAGCAAGACAAAGGCGCGACCTTCAGCGGAAATGACGCGGACGCGGCGGCTTACAG 360  
DB 301 AAGCAAGACAAAGGCGCGACCTTCAGCGGAAATGACGCGGACGCGGCGGCTTACAG 360  
QY 361 TTGCTGCGGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 420  
DB 361 TTGCTGCGGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 420  
QY 421 GAGTTGCAAGCGATCTGTAAGCAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
DB 421 GAGTTGCAAGCGATCTGTAAGCAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
QY 481 GACGTGTCGCGCTTCTCGCGCTGAAAGTCATGACGCGGATTCGATGTCGGAATTC 540  
DB 481 GACGTGTCGCGCTTCTCGCGCTGAAAGTCATGACGCGGATTCGATGTCGGAATTC 540  
QY 541 CTCGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600  
DB 541 CTCGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600  
QY 601 ATGACCTGATTCACCTGATTAATGCGGCTGCAAGCGGTCGACAAAGAAATCTTCGCTGAAG 660  
DB 601 ATGACCTGATTCACCTGATTAATGCGGCTGCAAGCGGTCGACAAAGAAATCTTCGCTGAAG 660  
QY 661 TCGCGGAAGATTCGACACGACGAAGAACTTCAGCGCGAAATTCGCGGCTGCGGCGG 720  
DB 661 TCGCGGAAGATTCGACACGACGAAGAACTTCAGCGCGAAATTCGCGGCTGCGGCGG 720  
QY 721 AAGGCGCGCTTACGCTTGTGAGCTTGTTCGAGACGAGAGATACGCGGACGCGGA 780  
DB 721 AAGGCGCGCTTACGCTTGTGAGCTTGTTCGAGACGAGAGATACGCGGACGCGGA 780  
QY 781 ATGTCGAATGTCGATCAACAAAGTTAGGCACTGACACTCCCTTACCGGACCTTC 840  
DB 781 ATGTCGAATGTCGATCAACAAAGTTAGGCACTGACACTCCCTTACCGGACCTTC 840  
QY 841 GAGTTGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900  
DB 841 GAGTTGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900  
QY 901 CCCTTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960  
DB 901 CCCTTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960  
QY 961 CGCATGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020  
DB 961 CGCATGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020  
QY 1021 AAGCGCTGAGACCGGCAACCGTTATGCGAATCTTCGCGGACCGCGATTCGCGGCTTC 1080  
DB 1021 AAGCGCTGAGACCGGCAACCGTTATGCGAATCTTCGCGGACCGCGATTCGCGGCTTC 1080





XX (BADI) BASF PLANT SCI GMBH.  
PA (MCKE/) MCKERSIE B.  
XX  
PI Mankin L:  
XX  
DR WPI: 2002-280939/32.  
XX  
PT New self-excising polynucleotides, useful for producing transgenic  
PT plants, removing transgenes from these plants or crops (e.g. food  
PT commodities), and restricting the distribution of transgenes within the  
PT environment.  
XX  
PS Example 2; Fig 8; 60pp; English.  
XX  
CC The invention describes an isolated excisable polynucleotide, which  
CC comprises a desired trait polynucleotide and a recombinase polynucleotide  
CC operably linked to a promoter, all flanked by a pair of directly oriented  
CC recombination sites where the recombinase activity is regulatable. The  
CC self-excising polynucleotide is useful for producing transgenic plants,  
CC particularly for removing all transgenic nucleic acid sequences that  
CC encode a gene product from the transgenic plant, thus restoring the  
CC original genetic configuration of the genome. The polynucleotide is also  
CC useful in methods for eliminating unwanted nucleic acids in agricultural  
CC food products and for preventing the escape of certain transgenic traits  
CC into the environment. This sequence represents the plasmid pBPS BW051  
CC T-DNA region used to test the self-excising DNA described in the  
CC invention.  
CC  
XX  
SQ Sequence 9880 BP; 2489 A; 2368 C; 2475 G; 2548 T; 0 other:  
  
Query Match 99.2%; Score 2015; DB 24; Length 9880;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 2021; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
  
QY 1 ATGGCACAAGGGGTTGTGACCGGGGTGATACGTAAGTTTCTGCTTACCTTGTATATA 60  
DB 4977 ATGGCACAAGGGGTTGTGACCGGGGTGATACGTAAGTTTCTGCTTACCTTGTATATA 5036  
QY 61 TAT 120  
DB 5037 TAT 5096  
QY 121 AAGAAT 180  
DB 5097 AAGAAT 5156  
QY 181 ATAACTTTCTAAT 240  
DB 5157 ATAACTTTCTAAT 5216  
QY 241 CGTCAGTCGGGCGGACCGCGAAGATAGACGTCAGCAAGCCGACACAGCAGTACGCC 300  
DB 5217 CGTCAGTCGGGCGGACCGCGAAGATAGACGTCAGCAAGCCGACACAGCAGTACGCC 5276  
QY 301 AACGAAGACAAAGCGCGGACCTTACGCGCAAGTCAGCGCGGCGGCGGCGGCTTACG 360  
DB 5277 AACGAAGACAAAGCGCGGACCTTACGCGCAAGTCAGCGCGGCGGCGGCGGCTTACG 5336  
QY 361 TTGCTGCGGCAATTTACGCAAGCGCGGCAAGTCAGCGCGGCGGCGGCGGCGGCGGCG 420  
DB 5337 TTGCTGCGGCAATTTACGCAAGCGCGGCAAGTCAGCGCGGCGGCGGCGGCGGCGGCG 5396  
QY 421 GAGTTGCAAGCAATCCGCAAGCAATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 480  
DB 5397 GAGTTGCAAGCAATCCGCAAGCAATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 5456  
QY 481 GAGCTGCGGCGCTTCTGCGCGCTGAAGTCATGAGCAGATTCGATGCTCGGAATTTG 540  
DB 5457 GAGCTGCGGCGCTTCTGCGCGCTGAAGTCATGAGCAGATTCGATGCTCGGAATTTG 5516  
QY 541 CTCGCCCTGGGCGTGACGATTTGTTCCACTCAGGAAGGCGCTTTCGCGGACGGAACGTC 600  
|||||

DB 5517 CTCGCCCTGGGCGTGACGATTTGTTCCACTCAGGAAGGCGCTTTCGCGGACGGAACGTC 5576  
QY 601 ATGGACCTGATTCACCGGATTTATGCGCTGACGCGTCGACAAAGATCTGCTGAAG 660  
DB 5577 ATGGACCTGATTCACCGGATTTATGCGCTGACGCGTCGACAAAGATCTGCTGAAG 5636  
QY 661 TCGGCGAAGATTCGACACGAGAAACCTTCAGCGGCAATTTGGCGGCTAGTCGCGGCG 720  
DB 5637 TCGGCGAAGATTCGACACGAGAAACCTTCAGCGGCAATTTGGCGGCTAGTCGCGGCG 5696  
QY 721 AAGCGCCTTACGCGCTTCGACGCTTTGTTGAGACGAAAGATTCAGCGCACGCGCGA 780  
DB 5697 AAGCGCCTTACGCGCTTCGACGCTTTGTTGAGACGAAAGATTCAGCGCACGCGCGA 5756  
QY 781 ATGCTCAATGTCGATCAACAAATAGCGGACGTCGACGCTTACCGGACCGCTTC 840  
DB 5757 ATGCTCAATGTCGATCAACAAATAGCGGACGTCGACGCTTACCGGACCGCTTC 5816  
QY 841 GAGTTGAGCGCGGACGTAATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
DB 5817 GAGTTGAGCGCGGACGTAATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5876  
QY 901 CCCTTCAAGCGCGGACGTAACCGGCAATTCACCGGCGACATCAGCGGCGCTTTGTAAG 960  
DB 5877 CCCTTCAAGCGCGGACGTAACCGGCAATTCACCGGCGACATCAGCGGCGCTTTGTAAG 5936  
QY 961 CGCATGAGCGGTCGACGCGGCTGCGACCGGCGGCGGAGAGATTTGGAGAAACACGCTTCA 1020  
DB 5937 CGCATGAGCGGTCGACGCGGCTGCGACCGGCGGCGGAGAGATTTGGAGAAACACGCTTCA 5996  
QY 1021 AGCGCTTGGGACCGCGGACCGGTAATGCAATCTTGGGACCGCGGCTTATGCGGCTTC 1080  
DB 5997 AGCGCTTGGGACCGCGGACCGGTAATGCAATCTTGGGACCGCGGCTTATGCGGCTTC 6056  
QY 1081 GCGCTGAGGTCATCTACAGAAAGACCGGACGCGGACCGGACGAGATTTGAGGCT 1140  
DB 6057 GCGCTGAGGTCATCTACAGAAAGACCGGACGCGGACCGGACGAGATTTGAGGCT 6116  
QY 1141 TACGCGAATTCAGCGGACCGGATCAGCTCGGCGGCGGCTGAGATTTGGGACCGGCTC 1200  
DB 6117 TACGCGAATTCAGCGGACCGGATCAGCTCGGCGGCGGCTGAGATTTGGGACCGGCTC 6176  
QY 1201 ATCAGGCGCGGTCAGTGTATGAGCTTCAGGCGTGTGAGACGCGGACGCGGCGGCAAG 1260  
DB 6177 ATCAGGCGCGGTCAGTGTATGAGCTTCAGGCGTGTGAGACGCGGACGCGGCGGCAAG 6236  
QY 1261 GGGCTTCCGCGGCGGCAAGCCATTCTGCGCCATGACAGAGCTGTACTGCGAGTGTGCG 6296  
DB 6237 GGGCTTCCGCGGCGGCAAGCCATTCTGCGCCATGACAGAGCTGTACTGCGAGTGTGCG 6296  
QY 1321 GCGCTCAATGATTCGAAGCGGCGGGAAGATCAGATCAAGGACCTTACCGGCTGCGGCG 1380  
DB 6297 GCGCTCAATGATTCGAAGCGGCGGGAAGATCAGATCAAGGACCTTACCGGCTGCGGCG 6356  
QY 1381 CGGAAGGTGTGTCGACCGCTGCGGACCTTGGGACGAGCAAGACGACGTCGACGATG 1440  
DB 6357 CGGAAGGTGTGTCGACCGCTGCGGACCTTGGGACGAGCAAGACGACGTCGACGATG 6416  
QY 1441 GCGGCACTCGACAGATTCGTTGCGGAACGATCTTCAACAAGATCAGGACCGCGGAAGGC 1500  
DB 6417 GCGGCACTCGACAGATTCGTTGCGGAACGATCTTCAACAAGATCAGGACCGCGGAAGGC 6476  
QY 1501 GAGCAAGAGAGCTTGGGCTTCTGTTGGGAAGCGCGGACGCTTCGGAACCTCAGTAGAG 1560  
DB 6477 GAGCAAGAGAGCTTGGGCTTCTGTTGGGAAGCGCGGACGCTTCGGAACCTCAGTAGAG 6536  
QY 1561 GCGCCTGAGAGAGCGGCGGAACGCGGCAACCTTTCGCGGAGCGCGGCGGCGGCGGCTTAC 1620  
DB 6537 GCGCCTGAGAGAGCGGCGGAACGCGGCAACCTTTCGCGGAGCGCGGCGGCGGCGGCTTAC 6596  
QY 1621 GCGCTTGAAGAGCTGTACGAAGACCGGCGGCGGAGAGCTTACGACGAGCCGTTGGCAGG 1680  
DB 6597 GCGCTTGAAGAGCTGTACGAAGACCGGCGGCGGAGAGCTTACGACGAGCCGTTGGCAGG 6656  
|||||

Qy	1681	AGACATTCGGGAAGCAAGCAGGAGCGTGACCTCCGGCAGCAAGGCGGAGAAGCGG	1740
Db	6657	AAGCATTCCGGAAGCAAGCAGGAGCGTGACCTCCGGCAGCAAGGCGGAGAAGCGG	6716
Qy	1741	CTTGCCGAACCTTAAGCGCGCGGAAGCCCGGAAGTTGCTCCCTTGACCAATGTTCCCGGA	1800
Db	6717	CTTGCCGAACCTTAAGCGCGCGGAAGCCCGGAAGTTGCTCCCTTGACCAATGTTCCCGGA	6776
Qy	1801	GAGCGCAGCCTACCCCGACCGCCCTTAAGTGTGTGGGGCGCGCGTCACTAGACAC	1860
Db	6777	GAGCGCAGCCTACCCCGACCGCCCTTAAGTGTGTGGGGCGCGCGTCACTAGACAC	6836
Qy	1861	AAGCGCGTGTTCGTGCGGCTCTTCGTAGACAAGATCGTTGTACGAAAGTCGACTAGGGG	1920
Db	6837	AAGCGCGTGTTCGTGCGGCTCTTCGTAGACAAGATCGTTGTACGAAAGTCGACTAGGGG	6896
Qy	1921	AGGGGGCAGGGAAGCGCCCATCGAAGAAGCGCGCTTGATCATCGTGGGCGAAGCGCGGAC	1980
Db	6897	AGGGGGCAGGGAAGCGCCCATCGAAGAAGCGCGCTTGATCATCGTGGGCGAAGCGCGGAC	6956
Qy	1981	GACGACGACGAAGACGACGCCCGACGAGCGGACGAGGAAGCGTAGCGGGCGTAG	2031
Db	6957	GACGACGACGAAGACGACGCCCGACGAGCGGACGAGGAAGCGTAGCGGGCGTAG	7007
RESULT 4			
ID	AAD04935	AAD04935 standard; DNA; 5711 BP.	
XX	AAD04935;		
AC	17-JUL-2001	(first entry)	
DT	C31-Int expression vector pRK65 for phiC31 integrase.		
DE	Gene trapping construct; conditional mutation; unidirectional inversion		
KW	recombinase recognition sequence; RRS; disruption cassette;		
KW	selection cassette; transgenic organism; expression vector pRK65;		
KW	phiC31 integrase; C31-Int mediated inversion; ds.		
XX	Chimeric - Bacteriophage phi-C31.		
OS	Chimeric - Cytomegalovirus.		
OS	Chimeric - Unidentified.		
XX	Key		
PH	Promoter	Location/Qualifiers	
FT		1..700	
FT		/*tag- a	
FT	Intron	/note- "Cytomegalovirus immediate early gene promoter"	
FT		700..970	
FT		/*tag- b	
FT		/note- "Hybrid Intron"	
FT	CDS	978..2819	
FT		/*tag- c	
FT	misc.feature	/product- "C31-Int, a phiC31 phage derived integrase"	
FT		2831..3020	
FT		/*tag- d	
FT		/note- "Polyadenylation sequence"	
XX	MO200129208-A1.		
PN	26-APR-2001.		
PD	16-OCT-2000; 2000WO-EPI0162.		
XX	16-OCT-1999; 99EP-0120592.		
XX	27-OCT-1999; 99US-0162016.		
XX	(ARTE-) ARTEMIS PHARM GMBH.		
PA	(FRAN-) FRANKFEN BIOTECHNOLOGIE AG.		
XX	Kuehn R, Von Melchner H, Altschmed J;		
XX			

[illegible]

QY 813 CTGCACCACTCCCTTACCGGACCTTCGAGTTCGAGCCGACGTAATCCGGTGTG 872  
 DB 1601 CTCGACCACTCCCTTACCGGACCTTCGAGTTCGAGCCGACGTAATCCGGTGTG 1660  
 QY 873 GCGTGAATCAAGACGACAAACACTTCCCTTCAAGCCGGGAGTCAAGCCGCATTCA 932  
 DB 1661 GCGTGAATCAAGACGACAAACACTTCCCTTCAAGCCGGGAGTCAAGCCGCATTCA 1720  
 QY 933 CCCGGGACGATCAAGGGGCTTTTGAAGCCGATGAGCCGTCAGCCGCGCCAGCCGGG 992  
 DB 1721 CCCGGGACGATCAAGGGGCTTTTGAAGCCGATGAGCCGTCAGCCGCGCCAGCCGGG 1780  
 QY 993 CGAGACGATTGGGAAGAAGACCGCTTCAAGCCCTGGGACCCGGCAACCTTATGCAAT 1052  
 DB 1781 CGAGACGATTGGGAAGAAGACCGCTTCAAGCCCTGGGACCCGGCAACCTTATGCAAT 1840  
 QY 1053 CTTTCGGGACCCCGCTATTTGGGGCTTCGCCCTGAGTGTATTCAGAAAGAACCCGA 1112  
 DB 1841 CTTTCGGGACCCCGCTATTTGGGGCTTCGCCCTGAGTGTATTCAGAAAGAACCCGA 1900  
 QY 1113 CGCGACGCGGACGACGAAGATTGAGGGTTACCGATTGAGGGGACCCGATCAGCGTCG 1172  
 DB 1901 CGCGACGCGGACGACGAAGATTGAGGGTTACCGATTGAGGGGACCCGATCAGCGTCG 1960  
 QY 1173 GCCGCTGAGCTTATTTGGGACCGCATATCAGCCCGCTGAGTGTATGAGCTTCAAGC 1232  
 DB 1961 GCCGCTGAGCTTATTTGGGACCGCATATCAGCCCGCTGAGTGTATGAGCTTCAAGC 2020  
 QY 1233 GTGCTTGGAGCGGAGGGGGCGGCAAGGGGCTTTCCGGGGGCAAGCCATTCTGTCCGC 1292  
 DB 2021 GTGCTTGGAGCGGAGGGGGCGGCAAGGGGCTTTCCGGGGGCAAGCCATTCTGTCCGC 2080  
 QY 1293 CATGACAGCTGATGAGCGGAGTGGGGCGGCTGATGATTCGAAAGCCGGGGAAGATC 1352  
 DB 2081 CATGACAGCTGATGAGCGGAGTGGGGCGGCTGATGATTCGAAAGCCGGGGAAGATC 2140  
 QY 1353 GATCAAGGACTCTTACCGCTGCTCGCGGAGAGGTGTTGACCCGTCACCTGAGCA 1412  
 DB 2141 GATCAAGGACTCTTACCGCTGCTCGCGGAGAGGTGTTGACCCGTCACCTGAGCA 2200  
 QY 1413 GCACGAAGGCACTGCAACGTCAGATGGCGGCACTGCAAGATTCGTTGCGGAACGAT 1472  
 DB 2201 GCACGAAGGCACTGCAACGTCAGATGGCGGCACTGCAAGATTCGTTGCGGAACGAT 2260  
 QY 1473 CTTCAACAGATCAGGAGCGCGGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1532  
 DB 2261 CTTCAACAGATCAGGAGCGCGGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2320  
 QY 1533 CGCCGAGCGCTTCGCAAGCTCACTGAGGCGCTGAGAAAGAGCGGCAAGCGGAGACT 1592  
 DB 2321 CGCCGAGCGCTTCGCAAGCTCACTGAGGCGCTGAGAAAGAGCGGCAAGCGGAGACT 2380  
 QY 1593 TGTTCGGGAGCGCGGACCCCTGGAAGCCCTTGAAGAGCTGTACGAAGACCCGCGGC 1652  
 DB 2381 TGTTCGGGAGCGCGGACCCCTGGAAGCCCTTGAAGAGCTGTACGAAGACCCGCGGC 2440  
 QY 1653 AGGAGCTTACGAG 1712  
 DB 2441 AGGAGCTTACGAG 2500  
 QY 1713 GCTTCGGGACGAGAGGCGGAGAGCGGCTTGCAGAACTTGAAGCCGCGGAGCCGGA 1772  
 DB 2501 GCTTCGGGACGAGAGGCGGAGAGCGGCTTGCAGAACTTGAAGCCGCGGAGCCGGA 2560  
 QY 1773 GTTCGCCCTTGAACATGTTCCCGGAAGAGCGGACGCTGACCCGAGCCCTTAAGTC 1832  
 DB 2561 GTTCGCCCTTGAACATGTTCCCGGAAGAGCGGACGCTGACCCGAGCCCTTAAGTC 2620  
 QY 1833 GTGGTGGGGGCGGCTGAG 1892  
 DB 2621 GTGGTGGGGGCGGCTGAG 2680  
 QY 1893 GATGCTTGTGACGAAGTGCATACGGGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1952

DB 2681 GATGCTTGTGACGAAGTGCATACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2740  
 QY 1953 TTGCATCAGCTGGGGGAGAGCGCCGACGACGACGACGACGACGACGACGACGACGAC 2012  
 DB 2741 TTGCATCAGCTGGGGGAGAGCGCCGACGACGACGACGACGACGACGACGACGACGAC 2800  
 QY 2013 GGAAGAGCTAGCGGCTAG 2031  
 DB 2801 GGAAGAGCTAGCGGCTAG 2819  
 RESULT 5  
 AAH74879  
 ID AAH74879 standard; DNA; 1842 BP.  
 AC  
 XX AAH74879:  
 DT 29-OCT-2001 (first entry)  
 XX  
 DE Nucleotide sequence of a wildtype bacteriophage thetaC31 integrase.  
 XX  
 KW Recombinase; genetic engineering; monogenic disorder; ADA deficiency;  
 KW cystic fibrosis; familial-hypercholesterolaemia; haemophilia;  
 KW chronic granulomatous disease; Duchenne's muscular dystrophy;  
 KW Fanconi's disease; anemia; sickle-cell anemia; Gaucher's disease;  
 KW Hunter's syndrome; X-linked severe combined immunodeficiency; SCID;  
 KW infectious disease; acquired disorder; tumour; cancer; integrase; ss.  
 OS  
 XX Bacteriophage thetaC31.  
 OS  
 PN WO200161049-A1.  
 XX  
 PD 23-AUG-2001.  
 XX  
 PF 16-FEB-2001; 2001WO-US05269.  
 XX  
 PR 18-FEB-2000; 2000US-0183759.  
 XX  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PA  
 PI Calos MP, Sclimenti CR:  
 XX  
 PI Identifying altered recombinase, involves hybridizing recombination  
 PT sites in cells having specific coding sequence, transformed with  
 PT altered recombinase gene, and isolating cells having product of the  
 PT sequence -  
 PS  
 XX Disclosure; Fig 4; 101pp; English.  
 XX  
 CC The specification describes a method for identifying altered  
 CC recombinases. The method comprises transforming cells with a first  
 CC plasmid comprising two recombination sites and a coding sequence of  
 CC interest, and a second plasmid encoding an altered recombinase; allowing  
 CC recombination of the two recombination sites using the altered  
 CC recombinase; screening and isolating transformed cells comprising the  
 CC product of the sequence of interest; and identifying the altered  
 CC recombinase. The altered recombinase is useful for site-specifically  
 CC integrating a polynucleotide sequence of interest in a genome of a cell.  
 CC It is also useful in genetic engineering of chromosomes of higher cells,  
 CC and for the generation of transgenic cells, tissues, plants and animals.  
 CC The altered recombinase is useful for treating monogenic disorders,  
 CC e.g. ADA deficiency, cystic fibrosis, familial-hypercholesterolaemia,  
 CC anemia, chronic granulomatous disease, Duchenne's muscular dystrophy,  
 CC haemophilia, Fanconi's disease, sickle-cell anemia, Gaucher's disease,  
 CC Hunter's syndrome and X-linked severe combined immunodeficiency (SCID),  
 CC infectious diseases including viral and bacterial infections, acquired  
 CC disorders including solid tumours and haematopoietic cancers such as  
 CC leukaemias and lymphomas, and other cancers. The present sequence  
 CC encodes a wild type bacteriophage thetaC31 integrase.  
 CC

SO Sequence 1842 BP; 387 A; 547 C; 616 G; 292 T; 0 other;

Query Match 88.4%; Score 1795; DB 22; Length 1842;  
Best Local Similarity 99.2%; Pred. No. 0;  
Matches 1804; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

213 GATGTGACGATGACGGGGTGTCTTACGACCGTCAATCGCGGAGCGGAGAAATACAGTGC 272  
24 GGTGGACACGACGGGGTGTCTTACGACCGTCAATCGCGGAGCGGAGAAATACAGTGC 83  
273 AGCAAGCCACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 332  
84 AGCAAGCCACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 143  
333 AGTGAACGCGACGCGGGCGCGGTTCAGTTCGTGCGGCAATTCAGCGAACGCGCGGAC 392  
144 AGTGAACGCGACGCGGGCGCGGTTCAGTTCGTGCGGCAATTCAGCGAACGCGCGGAC 203  
393 GTGCGGCTTGGGAGCGCGGAGCGCGGAGTTCGAAAGCATCTCTGACGAATTCGCGGC 452  
204 GTGCGGCTTGGGAGCGCGGAGCGCGGAGTTCGAAAGCATCTCTGACGAATTCGCGGC 263  
453 CGGCGGCTTCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 512  
264 CGGCGGCTTCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 323  
513 GGAAGGCTTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 572  
324 GGAAGGCTTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 383  
573 GGAAGGCTTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 632  
384 GGAAGGCTTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 443  
633 CGGCGGCTTCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 692  
444 CGGCGGCTTCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 503  
693 GCGCGAATTTGGGAGCGCGGAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 752  
504 GCGCGAATTTGGGAGCGCGGAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 563  
753 GACGAGAGAGATCAGCGGACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 812  
564 GACGAGAGAGATCAGCGGACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 623  
813 CTGACACATCCCTTACCGGACCTTTCAGATTCGAGCCGACGATTCGAGTGTGTGTGT 872  
624 CTGACACATCCCTTACCGGACCTTTCAGATTCGAGCCGACGATTCGAGTGTGTGTGT 683  
873 GCGGTGATCAGAGACGACAAACACCTTTCAGATTCGAGCCGACGATTCGAGTGTGTGT 932  
684 GCGGTGATCAGAGACGACAAACACCTTTCAGATTCGAGCCGACGATTCGAGTGTGTGT 743  
933 CCCGGGACGATCAGCGGGCTTGTAGCGCATGAGCGCATGAGCGCATGAGCGCATGAGCG 992  
744 CCCGGGACGATCAGCGGGCTTGTAGCGCATGAGCGCATGAGCGCATGAGCGCATGAGCG 803  
993 CGAGACGATTTGGAGAGAGACGCTTCAAGCCCTTGGAGCCGCGGACCGCTTATGCGAAT 1052  
804 CGAGACGATTTGGAGAGAGACGCTTCAAGCCCTTGGAGCCGCGGACCGCTTATGCGAAT 863  
1053 CCTTGGGAGCCGCTTATTTGGGAGCTTTCAGGCTGAGTGTATTCAGAGAGAGCGGGA 1112  
864 CCTTGGGAGCCGCTTATTTGGGAGCTTTCAGGCTGAGTGTATTCAGAGAGAGCGGGA 923  
1113 CGGACGCGCGACGAGAGATTTAGGCTTACCGCATTCAGCGCGACCGCATCAGCTCCG 1172  
924 CGGACGCGCGACGAGAGATTTAGGCTTACCGCATTCAGCGCGACCGCATCAGCTCCG 983  
1173 GCGGCTGAGGCTTATTTGGGAGCGGATTCAGGAGCGGCTTATGAGTGTATGAGTGTATG 1232  
984 GCGGCTGAGGCTTATTTGGGAGCGGATTCAGGAGCGGCTTATGAGTGTATGAGTGTATG 1043

QY 1233 GTGTTGAGCGGACGAGGCGCGGCGGCAAGGCGCTTTCGCGGAGCAACCATTCCTGCGC 1292  
DB 1044 GTGTTGAGCGGACGAGGCGCGGCGGCAAGGCGCTTTCGCGGAGCAACCATTCCTGCGC 1103  
QY 1293 CATGAGCAAGCTTACTGAGAGTGTGGCGCGGCTCATGACTTTCGAGGAGGAGGAGGAG 1352  
DB 1104 CATGAGCAAGCTTACTGAGAGTGTGGCGCGGCTCATGACTTTCGAGGAGGAGGAGGAG 1163  
QY 1353 GATCAAGGACTCTTACCGCTGCCGTCGCGGAGAGTGTGACGCGCTCCGACCTGGGCA 1412  
DB 1164 GATCAAGGACTCTTACCGCTGCCGTCGCGGAGAGTGTGACGCGCTCCGACCTGGGCA 1223  
QY 1413 GCACGAAGCAGCTGCAAGCTGACATGAGCGGACCTTCGCAAGTTCGTTGGGAAAGC 1472  
DB 1224 GCACGAAGCAGCTGCAAGCTGACATGAGCGGACCTTCGCAAGTTCGTTGGGAAAGC 1283  
QY 1473 CTTCACAAAGATCAGGACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1532  
DB 1284 CTTCACAAAGATCAGGACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1343  
QY 1533 CGCGGAGCGCTTCCGCAAGCTTACTGAGCGGCTTTCAGAAAGCGGCGGCAACCT 1592  
DB 1344 CGCGGAGCGCTTCCGCAAGCTTACTGAGCGGCTTTCAGAAAGCGGCGGCAACCT 1403  
QY 1593 TGTTCGAGAGCGCGCGGACGCGGCTTGAAGAGCTTACGAAAGCGGCGGCGG 1652  
DB 1404 TGTTCGAGAGCGCGCGGACGCGGCTTGAAGAGCTTACGAAAGCGGCGGCGG 1463  
QY 1653 AGGACTTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1712  
DB 1464 AGGACTTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1523  
QY 1713 GCTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1772  
DB 1524 GCTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1583  
QY 1773 GTTCCCTTTCAGCAAGTGTTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1832  
DB 1584 GTTCCCTTTCAGCAAGTGTTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1643  
QY 1833 GTGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1892  
DB 1644 GTGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1703  
QY 1893 GATCGTTTCAGCAAGTGTTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1952  
DB 1704 GATCGTTTCAGCAAGTGTTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1763  
QY 1953 TTTCGATCAGTGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2012  
DB 1764 TTTCGATCAGTGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1823  
QY 2013 GGAAGAGCTAGCGGCTAG 2031  
DB 1824 GGAAGAGCTAGCGGCTAG 1842

RESULT 6  
AA06904  
ID AA06904 standard; DNA: 3401 BP.  
AC AA06904;  
XX  
XX 06-MAR-1991 (first entry)  
XX  
XX Sequence encoding site specific Integrating function of actinomycete  
XX DE phage phi C31.  
XX  
XX Antibiotic production; isovaleryl; spiramycin; ds.  
XX  
XX Page phi C31.  
XX

PN EP403173-A.  
XX 19-DEC-1990.  
XX 08-JUN-1990: 90EP-0306260.  
XX 12-JUN-1989: 89US-0364959.  
XX (ELIL ) ELI LILLY & CO.  
XX Cox KL, Richardson MA, Kuhstoss SA, Schoner BE, Seno ET, Rao RN:  
XX WPI: 1990-377895/51.  
DR Use of site-specific integrating function of phage 0C31 - for  
XX increased prodn. of antibiotics and prodn. of hybrid antibiotics  
XX in actinomycetes  
XX  
XX Disclosure: Page 4-6: 46pp; English.  
XX  
XX Sequence may be incorporated into a plasmid with an antibiotic  
XX biosynthetic gene (eg. Isovaleryl or spiramycin) and used to  
XX transform a microorganism with site specific integration.  
XX Plasmids PKC796, POU243, PSK50 and PSK51 and actinomycetes  
XX transformed by them, are specifically claimed.  
XX  
XX Sequence 3401 BP: 660 A; 1011 C; 1135 G; 594 T; 1 other:  
SO  
Query Match 88.3%; Score 1793.4; DB 11; Length 3401;  
Best Local Similarity 99.1%; Pred. No. 0;  
Matches 1803; Conservative 0; Mismatches 16; Indels 0; Gaps 0;  
OY 213 GATGTCAGTACGGGGGCTTACGACCGTCACTGTCGCGAGCGGAGAAATGACATGTC 272  
DB 532 GTTGACACATGCGGGGCTTACGACCGTCACTGTCGCGAGCGGAGAAATTCGAGCGC 591  
OY 273 ACCAAGCCAGCGACACACGCTAGCGCCACGAAAGAGAGGCGCGACCTTCAGCGCA 332  
DB 592 ACCAAGCCAGCGACACACGCTAGCGCCACGAAAGAGAGGCGCGCGACCTTCAGCGCA 651  
OY 333 ACTCGAGCGCGAGCGGGGCGGCTTCAAGTTCGCGGCAATTCAGCGAAAGCGCGGCGC 392  
DB 652 ACTCGAGCGCGAGCGGGGCGGCTTCAAGTTCGCGGCAATTCAGCGAAAGCGCGGCGC 711  
OY 393 GTCGGGCTTCGGAGCGGGGCGGCGCGGAGTTCGAAAGCATCTGAAAGAAATGCGCGCGC 452  
DB 712 GTCGGGCTTCGGAGCGGGGCGGCGCGGAGTTCGAAAGCATCTGAAAGAAATGCGCGCGC 771  
OY 453 CGGGGCGCTCAACATGATCTGTATGATGCTGTCGCGCTTCGCGCTGAGGTCAAT 512  
DB 772 CGGGGCGCTCAACATGATCTGTATGATGCTGTCGCGCTTCGCGCTGAGGTCAAT 831  
OY 513 GGACGCGATTCGATGTTGTCTGGAATTCGCGCTTCGCGCTGAGCATTTTCCACTCA 572  
DB 832 GGACGCGATTCGATGTTGTCTGGAATTCGCGCTTCGCGCTGAGCATTTTCCACTCA 891  
OY 573 GGAAGGCGCTTCGCGCGAGGAAGCATGACCTGATTCACCTGATTAATGCGGCTCGA 632  
DB 892 GGAAGGCGCTTCGCGCGAGGAAGCATGACCTGATTCACCTGATTAATGCGGCTCGA 951  
OY 633 CGCGTCGCAAAAGATCTTCGCTGAAATCGCGGAGAAATTCGACACGAAACCTTCA 692  
DB 952 CGCGTCGCAAAAGATCTTCGCTGAAATCGCGGAGAAATTCGACACGAAACCTTCA 1011  
OY 693 GCGCGAATTCGGGCGGCTGACGTCGCGGAGAAAGCGCTTCAGCGCTTCGAGCTTTGCGA 752  
DB 1012 GCGCGAATTCGGGCGGCTGACGTCGCGGAGAAAGCGCTTCAGCGCTTCGAGCTTTGCGA 1071  
OY 753 GACGAAGGATTCACGCGCAAGCGGAGATTCGATGCTGATCAACGAAGTTACGCA 812  
DB 1072 GACGAAGGATTCACGCGCAAGCGGAGATTCGATGCTGATCAACGAAGTTACGCA 1131  
OY 813 CTCGACCACTCCCTTACCGGACCTTCGATTCGAGCCGAGCATTAATCCGGTGTGTG 872

DB 1132 CTCGACCACTCCCTTACCGGACCTTCGATTCGAGCCGAGCATATCGGTGTGTG 1191  
OY 873 GCGTGAGATCAGAGCGACCAACACCTTCCTTCAGCGCGGAGATCAAGCCCATTC 932  
DB 1192 GCGTGAGATCAGAGCGACCAACACCTTCCTTCAGCGCGGAGATCAAGCCCATTC 1251  
OY 933 CCGGGGACATCAGCGGGGCTTTGTAAGCATGAGCGGTGAGCGGCGGAGCGG 992  
DB 1252 CCGGGGACATCAGCGGGGCTTTGTAAGCATGAGCGGTGAGCGGCGGAGCGG 1311  
OY 993 CGAGCATTCGAGGAAGAGACCGCTTCAGCGCGCTGAGACCGCGCAACCTTATGGAAT 1052  
DB 1312 CGAGCATTCGAGGAAGAGACCGCTTCAGCGCGCTGAGACCGCGCAACCTTATGGAAT 1371  
OY 1053 CCTTCGGAGCGCGCTTATTCGCGGCTTCGCGCTGAGGTATTCAGAGAGCGCGA 1112  
DB 1372 CCTTCGGAGCGCGCTTATTCGCGGCTTCGCGCTGAGGTATTCAGAGAGCGCGA 1431  
OY 1113 CGGACGCGGACGACGAAAGATGAGGTTACGCAATTCAGCGCGACCGGCTCAGCGTCG 1172  
DB 1432 CGGACGCGGACGACGAAAGATGAGGTTACGCAATTCAGCGCGACCGGCTCAGCGTCG 1491  
OY 1173 GCGGTCGAGCTTGAATTCGCGGACCGATCATGAGCGCGCTGAGGTATGAGCTCAGGC 1232  
DB 1492 GCGGTCGAGCTTGAATTCGCGGACCGATCATGAGCGCGCTGAGGTATGAGCTCAGGC 1551  
OY 1233 GTGCTTGAGCGGAGGCGGCGGCGCAAGGCGCTTCGCGGCGCAACCTTCTCCG 1292  
DB 1552 GTGCTTGAGCGGAGGCGGCGGCGCAAGGCGCTTCGCGGCGCAACCTTCTCCG 1611  
OY 1293 CATGGAAGGCTTACTGCGAGTGTGGCGCGCATGATCTTCGAAGCGCGGAGAAATC 1352  
DB 1612 CATGGAAGGCTTACTGCGAGTGTGGCGCGCATGATCTTCGAAGCGCGGAGAAATC 1671  
OY 1353 GATCAAGGACTTCTTACCGCTGCGCGCGCGGAGGATGTGTGACCGCTCCACCTGGCA 1412  
DB 1672 GATCAAGGACTTCTTACCGCTGCGCGCGCGGAGGATGTGTGACCGCTCCACCTGGCA 1731  
OY 1413 GCAGGAAGGACGCTGCAACGCTCAAGCATGCGCGCATCTGACAAATTTGTCGGAAGCAT 1472  
DB 1732 GCAGGAAGGACGCTGCAACGCTCAAGCATGCGCGCATCTGACAAATTTGTCGGAAGCAT 1791  
OY 1473 CTTCAACAGATCAGGACGCGCGAGCGCGAGGAGAGAGACCTTGTGTGGGAAGC 1532  
DB 1792 CTTCAACAGATCAGGACGCGCGAGCGCGAGGAGAGAGACCTTGTGTGGGAAGC 1851  
OY 1533 CGCCGAGCGCTTCGCAAGCTCACTGAGCGCGCTGAGAAAGAGCGCGGAAAGCGGAACT 1592  
DB 1852 CGCCGAGCGCTTCGCAAGCTCACTGAGCGCGCTGAGAAAGAGCGCGGAAAGCGGAACT 1911  
OY 1593 TGTGTGGAGAGCGCGCGCGCGCTTGAAGCGCTTGAAGAGCTTGAAGAGAGCGCGGCG 1652  
DB 1912 TGTGTGGAGAGCGCGCGCGCGCTTGAAGCGCTTGAAGAGAGCTTGAAGAGAGCGCGGCG 1971  
OY 1653 AGGAGCTTACGAGGAGCGCGCTTGAAGAGAGAGCTTGAAGAGAGAGAGAGAGAGAGAG 1712  
DB 1972 AGGAGCTTACGAGGAGCGCGCTTGAAGAGAGAGAGCTTGAAGAGAGAGAGAGAGAGAG 2031  
OY 1713 GCTCGGAGCAAGGAGGCGGAGAGAGCGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 1772  
DB 2032 GCTCGGAGCAAGGAGGCGGAGAGAGAGCGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 2091  
OY 1773 GTTGCGCTTGAACATGTTTCCCGGAGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAG 1832  
DB 2092 GTTGCGCTTGAACATGTTTCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2151  
OY 1833 GTGCTGGGAGCGCGCTCAGTACAGCAAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1892  
DB 2152 GTGCTGGGAGCGCGCTCAGTACAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2211  
OY 1893 GATCGTTGTCAGAGATCGATCAGGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1952



```

|||||
Db 924 CGGACGCCGACGACGAAGATTGAGGGTTACCCGATTACAGGGCAGCCGAGTACAGGCTCCG 983
Qy 1173 GCCGGTGAGCTTATTTCGGGACCGATCATCGAGCCCGCTGAGTGTATGAGCTTCAAGCC 1232
Db 984 GCCCGGTGAGCTTATTTCGGGACCGATCATCGAGCCCGCTGAGTGTATGAGCTTCAAGCC 1043
Qy 1233 GTGGTTGAGGAGCGGGGGGGCGGCAAGGGGCTTCCCGGGGGGAGCAATTCGTCCGC 1232
Db 1044 GTGGTTGAGGAGCGGGGGGGCGGCAAGGGGCTTCCCGGGGGGAGCAATTCGTCCGC 1103
Qy 1293 CATGAGACAGCTGTACTGCGAGTGTGGCGCGTTCATGACTTTCGAAAGCGCGGGGAAGATC 1352
Db 1104 CATGAGACAGCTGTACTGCGAGTGTGGCGCGTTCATGACTTTCGAAAGCGCGGGGAAGATC 1163
Qy 1353 GATCAAGAGCTTCTACCGCTGCCCTGCCCGGAAGGTGTGACCCCTCCGACCTGGGCA 1412
Db 1164 GATCAAGAGCTTCTACCGCTGCCCTGCCCGGAAGGTGTGACCCCTCCGACCTGGGCA 1223
Qy 1413 GCACGAGGACGTCGCAACGTCACATGTGGCGGCACTGCAACAGTTGTTGGGGAAGCAT 1472
Db 1224 GCACGAGGACGTCGCAACGTCACATGTGGCGGCACTGCAACAGTTGTTGGGGAAGCAT 1283
Qy 1473 CTTCAACAGATCAGCGACGCGCGGAAGCGAGCAAGACGTTGGCGCTTCTGTGGGAAGC 1532
Db 1284 CTTCAACAGATCAGCGACGCGCGGAAGCGAGCAAGACGTTGGCGCTTCTGTGGGAAGC 1343
Qy 1533 CGCGCCGACGCTTGCGGCAAGCTCACTGAGGGCGCTGAGAGAGCGCGGCAAGACCT 1552
Db 1344 CGCGCCGACGCTTGCGGCAAGCTCACTGAGGGCGCTGAGAGAGCGCGGCAAGACCT 1403
Qy 1593 TGTTCGCGAGCGCGCGACGCGCGCTGAAGCGCGCTGAAGAGCTGTACGAAGCGCGCGGC 1652
Db 1404 TGTTCGCGAGCGCGCGACGCGCGCTGAAGCGCGCTGTACGAAGAGCGCGCGGC 1463
Qy 1653 AGGAGCTTACGAGCGAGCCCGTTGGCAGGAAGACACTTCCGGAAGCAGAGCAGCGCTGAC 1712
Db 1464 AGGAGCTTACGAGCGAGCCCGTTGGCAGGAAGACACTTCCGGAAGCAGAGCAGCGCTGAC 1523
Qy 1713 GCTTCGGGACGACGAGGGCGGGAAGGGCGCTTGGCGGAACCTTGAACCGCGCGGAGCCGGA 1772
Db 1524 GCTTCGGGACGACGAGGGCGGGAAGGGCGCTTGGCGGAACCTTGAACCGCGCGGAGCCGGA 1583
Qy 1773 GTTCGCCCTTGACCAATGTGTTCCCGGAAGCGCGAGCGCTGACCCGCGCCTTAAGTC 1832
Db 1584 GCTTCGCCCTTGACCAATGTGTTCCCGGAAGCGCGAGCGCTGACCCGCGCCTTAAGTC 1643
Qy 1833 GTGGTGGGGGGCGCGCTCAGTAGAGCAGCAAGCGCGTGTCTCGGGCTCTTCTAGACAA 1892
Db 1644 GTGGTGGGGGGCGCGCTCAGTAGAGCAGCAAGCGCATGTCTCGGGCTCTTCTAGACAA 1703
Qy 1893 GATCGTTGTACGAAAGTCGACTAGGGGCGAGGGGCGGAGGAGCGCCCATTCGGAAGCGCGC 1952
Db 1704 GATCGTTGTACGAAAGTCGACTAGGGGCGAGGGGCGGAGGAGCGCCCATTCGGAAGCGCGC 1763
Qy 1953 TTTCGATCAGCTGTGGGCGAAGCGCGCGAGCGAGCAGCAAGACGCGCGCGAGCGGAC 2012
Db 1764 TTTCGATCAGCTGTGGGCGAAGCGCGCGAGCGAGCAGCAAGACGCGCGCGAGCGGAC 1823
Qy 2013 GGAAGACGTAGCGGCGTAG 2031
Db 1824 GCAAGACGTAGCGGCGTAG 1842

```

```

RESULT 8
AAH74881
ID AAH74881 standard; DNA; 1839 BP.
XX AAH74881:
XX
XX 29-OCT-2001 (first entry)
XX Nucleotide sequence of altered recombinase 5C1.
DE

```

```

XX
KW Recombinase; genetic engineering; monogenic disorder; ADA deficiency;
KW cystic fibrosis; familial hypercholesterolemia; haemophilia;
KW chronic granulomatous disease; Duchenne's muscular dystrophy;
KW Fanconi's disease; anemia; sickle-cell anemia; Gaucher's disease;
KW Hunter's syndrome; X-linked severe combined immunodeficiency; SCID;
KW infectious disease; acquired disorder; tumour; cancer; ss.
XX
OS Synthetic.
OS Bacteriophage.
XX
FH Key
FT CDS
FT
FT /tag= a
FT /product= "altered recombinase 5C1"
FT /note= "no termination codon given"
XX
PD W0200161049-A1.
PD 23-AUG-2001.
XX
PF 16-FEB-2001; 2001WO-US05269.
XX
PR 18-FEB-2000; 2000US-0183759.
XX
PA (STRD ) UNIV LELAND STANFORD JUNIOR.
XX
PI Calos MP, Scilment CR;
XX
XX WPI: 2001-522610/57.
XX P-PSDB: AAG63848.
XX
PT Identifying altered recombinase, involves hybridizing recombination
PT sites in cells having specific coding sequence, transformed with
PT altered recombinase gene, and isolating cells having product of the
PT sequence -
XX
PS Disclosure; Fig 8; 101pp; English.
XX
CC The specification describes a method for identifying altered
CC recombinases. The method comprises transforming cells with a first
CC plasmid comprising two recombination sites and a coding sequence of
CC interest, and a second plasmid encoding an altered recombinase; allowing
CC recombination of the two recombination sites using the altered
CC recombinase; screening and isolating transformed cells comprising the
CC product of the sequence of interest; and identifying the altered
CC recombinase. The altered recombinase is useful for site-specifically
CC integrating a polynucleotide sequence of interest in a genome of a cell.
CC It is also useful in genetic engineering of chromosomes of higher cells,
CC and for the generation of transgenic cells, tissues, plants and animals.
CC The altered recombinase is useful for treating monogenic disorders,
CC e.g. ADA deficiency, cystic fibrosis, familial hypercholesterolemia,
CC anemia, chronic granulomatous disease, Duchenne's muscular dystrophy,
CC haemophilia, Fanconi's disease, sickle-cell anemia, Gaucher's disease,
CC Hunter's syndrome and X-linked severe combined immunodeficiency (SCID),
CC infectious diseases including viral and bacterial infections, acquired
CC disorders including solid tumours and haematopoietic cancers such as
CC leukaemias and lymphomas, and other cancers. The present sequence
CC encodes an altered recombinase of the invention.
XX
SQ Sequence 1839 BP; 388 A; 547 C; 613 G; 291 T; 0 other;

```

```

Query Match 87.9%; Score 1785.6; DB 22; Length 1839;
Best Local Similarity 99.0%; Pred. No. 0; Mismatches 19; Indels 0; Gaps 0;
Matches 1797; Conservative 0;

```

```

Qy 213 GATGTGACGATACGGGGTGTGCTTACGACCGTCACTGCGCGGCGGAGAAATAGTACGTGC 272
Db 24 GTGTGACACGTACGGGGTGTGCTTACGACCGTCACTGCGCGGCGGAGAAATTCAGGCGC 83
Qy 273 AGCAAGCCCGACGACAGCGGTAGCGCGCAACGAGAGAGGGCGGCGACCTTACCGCGGA 332
Db 84 AGCAAGCCCGACGACAGCGGTAGCGCGCAACGAGAGAGGGCGGCGACCTTACCGCGGA 143

```



QY	333	AGTCAGACCGGAGCGGGGGCCGGTTACAGTTCGTGGGCAATTTCAGCGAAGCCCGGCGAC	392
Db	144	AGTCAGACCGGAGCGGGGGCCGGTTACAGTTCGTGGGCAATTTCAGCGAAGCCCGGCGAC	203
QY	393	GTCGGGGTTCCGGACCGCGGAGCGCCCGGAGTTCGAAAGCATCTCCGAAAGCATTCGCGC	452
Db	204	GTCGGGGTTCCGGACCGCGGAGCGCCCGGAGTTCGAAAGCATCTCCGAAAGCATTCGCGC	263
QY	453	CGGGCGGCTTCACATGATGATTTGTATATAGGTGTGCGGCTTTCGGCGCTGAAGGTAT	512
Db	264	CGGGCGGCTTCACATGATGATTTGTATATAGGTGTGCGGCTTTCGGCGCTGAAGGTAT	323
QY	513	GGACGGGATTCGATTTGTCTCGGAATTCCTCGCCCTGGGCGGTGACGATTTGTTCACCTCA	572
Db	324	GGACGGGATTCGATTTGTCTCGGAATTCCTCGCCCTGGGCGGTGACGATTTGTTCACCTCA	383
QY	573	GGAAGGCTCTTCGGGACGGGAAACGTCATGAGCTGATTCACCTGATTAATGCGGCTCGA	632
Db	384	GGAAGGCTCTTCGGGACGGGAAACGTCATGAGCTGATTCACCTGATTAATGCGGCTCGA	443
QY	633	CGCGTGGCCAAAGAATCTGTGCTGAATGTGGGGGAAGTTCTGACACGAAGAACTTCA	692
Db	444	CGCGTGGCCAAAGAATCTGTGCTGAATGTGGGGGAAGTTCTGACACGAAGAACTTCA	503
QY	693	CGCGCAATTTGGCGGGTACGTTCGGGGGGAAGCCGCTTACGGCTTCAGCTTATTTTCGA	752
Db	504	CGCGCAATTTGGCGGGTACGTTCGGGGGGAAGCCGCTTACGGCTTCAGCTTATTTTCGA	563
QY	753	GACGAAGGAGTACAGCGGCACAGGCCCAATGTCGTAATGTCGTATCAACAAGTTAGCGCA	812
Db	564	GACGAAGGAGTACAGCGGCACAGGCCCAATGTCGTAATGTCGTATCAACAAGTTAGCGCA	623
QY	813	CTTCGACCACTCCCTTTACCGGACCCCTTCGATTCGAGCTCGACCGCAGTATCCGGGTGGTG	872
Db	624	CTTCGACCACTCCCTTTACCGGACCCCTTCGATTCGAGCTCGACCGCAGTATCCGGGTGGTG	683
QY	873	GGGTAGATTCAAAGACCGCAACAACCTTCCCTTCAGCCGGGCAATCAAGCCGCATTTCA	932
Db	684	GGGTAGATTCAAAGACCGCAACAACCTTCCCTTCAGCCGGGCAATCAAGCCGCATTTCA	743
QY	933	CCCGGGCAGCATCAACGGGGCTTTGTAGAGCGCATGAGAGCGTGAAGCCGTCGAGCCCGGGG	992
Db	744	CCCGGGCAGCATCAACGGGGCTTTGTAGAGCGCATGAGAGCGTGAAGCCGTCGAGCCCGGGG	803
QY	993	CGAGACGATTTGGGAAGAAGACCCGCTTCAAGCGCCTGGGACCCGGCAACCGTTATGCGAT	1052
Db	804	CGAGACGATTTGGGAAGAAGACCCGCTTCAAGCGCCTGGGACCCGGCAACCGTTATGCGAT	863
QY	1053	CGTTTCGGGACCCGGGTATTGCGGGGCTTCGCGGCTGAGAGGTATCTTCACAAAGAAACCGGA	1112
Db	864	CGTTTCGGGACCCGGGTATTGCGGGGCTTCGCGGCTGAGAGGTATCTTCACAAAGAAACCGGA	923
QY	1113	CGGCACGCCGACCAACGAAGATTGAGGTTAACCGCATTCAGCGCGACCCGATAGCCTCGG	1172
Db	924	CGGCACGCCGACCAACGAAGATTGAGGTTAACCGCATTCAGCGCGACCCGATAGCCTCGG	983
QY	1173	GCGCGTGGAGCTTGAATTTGGGAGCCGAAATCGAGCCCGCTGAGTGGTATGACCTTCAGGC	1232
Db	984	GCGCGTGGAGCTTGAATTTGGGAGCCGAAATCGAGCCCGCTGAGTGGTATGAGCTTCAGGC	1043
QY	1233	GTCGTTGGAGCGGACGGGGGCGCGCAAGGAGGCTTTCGCGGGGCAAGCATTTCTGTCCG	1292
Db	1044	GTCGTTGGAGCGGACGGGGGCGCGCAAGGAGGCTTTCGCGGGGCAAGCATTTCTGTCCG	1103
QY	1293	CATGGCAACACTGATCTGCGAGATGTGTGGCCCGCATGACTTCGAAAGCGCGGGGAAGATC	1352
Db	1104	CATGGCAACACTGATCTGCGAGATGTGTGGCCCGCATGACTTCGAAAGCGCGGGGAAGATC	1163
QY	1353	GATCAAGAGATCTTACCGGCGCCGTGCGCGGAAGGAGTTCGACCCGTCGCAACTGTGGGCA	1412
Db	1164	GATCAAGAGATCTTACCGGCGCCGTGCGCGGAAGGAGTTCGACCCGTCGCAACTGTGGGCA	1223

QY	1413	GCACGAAGGGCACGTGCACACGTACATGCGCGCATCGACAAAGTTCTGTGGGAAACGAT	1472
Db	1224	GCACGAAGGGCACGTGCACACGTACATGCGCGCATCGACAAAGTTCTGTGGGAAACGAT	1293
QY	1473	CTTTCACACAGATCAGGCACCGCGGAAGGCGACGAAGAGACGTTGGCGCTTCTGTGGGAAGC	1532
Db	1284	CTTTCACACAAATCATGCGCACCCCGAAGGCGGACGAAGAGACGTTGGCGCTTCTGTGGGAAGC	1343
QY	1533	CGCCCGACGGCTTGGGCAAGCTCAGTGTAGGGGCGCTGAGAAAGACGGGCGAAGCGGAAACCT	1592
Db	1344	CGCCCGACGGCTTGGGCAAGCTCAGTGTAGGGGCGCTGAGAAAGACGGGCGAAGCGGAAACCT	1403
QY	1593	TGTTGCGGACCGCGCCGACGCCCTTGAACGGCCTTGAAGACCTGTACGAAGACCGCGGCGC	1652
Db	1404	TGTTGCGGACCGCGCCGACGCCCTTGAAGACCTGTACGAAGACCTGTACGAAGACCGCGGCGC	1463
QY	1653	AGGAGCTTACGACGAGGACCCGTTGGCAGGAAGCACCTTCGGAGCAACAGGCGACGCTGAC	1712
Db	1464	AGGAGCGGTACGACGAGACCCGTTGGCAGGAAGCACCTTCGGAGCAACAGGCGACGCTGAC	1523
QY	1713	GCCTCGGCGACGAAGGGGCGGAAAGAGGGGCTTGGCGAAACTTGAAGCGCGCGCAAGCCCGGA	1772
Db	1524	GCCTCGGCGACGAAGGGGCGGAAAGAGGGGCTTGGCGAAACTTGAAGCGCGCGCAAGCCCGGA	1583
QY	1773	GTTGCCCCCTTGACCAATGGTTCCTCCGAAAGACGCGCTGACCCGACCGGCCCTTAAGTC	1832
Db	1584	GCTTCCCTTGAACCAATGGTTCCTCCGAAAGACGCGCTGACCCGACCGGCCCTTAAGTC	1643
QY	1833	GTTGGTGGGGGCGGCGCTCAGTAGACGACAAGGCGGTTGCTGTGGGCTCTTGTGTACACA	1892
Db	1644	GTTGGTGGGGGCGGCGCTCAGTAGACGACAAGGCGGTTGCTGTGGGCTCTTGTGTACACA	1703
QY	1893	GATCGTTGTACGAGAGTGCAGTACGGGGCAGGGGGCAGGGAGCGCCCATCGAGAAAGCGCGC	1952
Db	1704	GATCGTTGTACGAGAGTGCAGTACGGGGCAGGGGGCAGGGAGCGCCCATCGAGAAAGCGCGC	1763
QY	1953	TTCGATCAGCTGGGGCGCAACCGCGCCGACCCAGCAGCAGCAAGAGACGCCAGGACGGCAC	2012
Db	1764	TTCGATCAGCTGGGGCGCAACCGCGCCGACCCAGCAGCAGCAAGAGACGCCAGGACGGCAC	1823
QY	2013	GGAGAGCGTAGCGCGG 2028	
Db	1824	GGAGAGCGTAGCGCGG 1839	
RESULT 9			
AAH74882			
ID	AAH74882 standard; DNA; 1833 BP.		
XX			
AC	AAH74882;		
XX			
DT	29-OCT-2001 (first entry)		
XX			
DE	Nucleotide sequence of altered recombinase 1C1.		
XX			
KW	Recombinase; genetic engineering; monogenic disorder; ADA deficiency;		
KW	cystic fibrosis; familial-hypercholesterolemia; haemophilia;		
KW	chronic granulomatous disease; Duchenne's muscular dystrophy;		
KW	Fanconi's disease; anemia; sickle-cell anemia; Gaucher's disease;		
KW	Hunter's syndrome; X-linked severe combined immunodeficiency; SCID;		
KW	Infectious disease; acquired disorder; tumour; cancer; ss.		
XX			
OS	Synthetic.		
XX			
OS	Bacteriophage.		
XX			
FT	Key		
FT	Location/Qualifiers		
FT	1..1833		
FT	/*tag= a		
XX	/product= "altered recombinase 1C1"		
PN	MO200161049-A1.		
XX			
DD	23-AUG-2001		



[illegible]

OY	633	CGCGTCGCAAAAGAATCTTCGCTGAAGTGGCGGAAGATTCTCGACAGAAACCTTCA	692
Db	444	CGCGTCGCAAAAGAATCTTCGCTGAAGTGGCGGAAGATTCTCGACAGAAACCTTCA	503
OY	693	GGCGGAATTTGGGGGGGATACGGCGGGGGAAGGGCGCTTACGGCTTCGACTTGTTCGGA	752
Db	504	GGCGGAATTTGGGGGGGATACGGCGGGGGAAGGGCGCTTACGGCTTCGACTTGTTCGGA	563
OY	753	GACGAAAGAGATTCACGGCGCAACGGCCGAATGTCAATGTCTCATCAACAAAGTTAGCCA	812
Db	564	GACGAAAGAGATTCACGGCGCAACGGCCGAATGTCTCATCAACAAAGTTAGCCA	623
OY	813	CTCGACACACTCCCTTACCGGACCTTCGATTCGAGTCCAGCCCGAGATTCGGTGGTG	872
Db	624	CTCGACACACTCCCTTACCGGACCTTCGATTCGAGTCCAGCCCGAGATTCGGTGGTG	683
OY	873	GGGTGAGATTCAAAGCGCAAAACCCCTTCGACGGGGGAGTCAAGCCGCATTTCA	932
Db	684	GGGTGAGATTCAAAGCGCAAAACCCCTTCGACGGGGGAGTCAAGCCGCATTTCA	743
OY	933	CCCCGGCGAGCTCACGGGGGCTTTGTAAAGCCATGAGAGCTGAGCGGCTCCGACCCGGG	992
Db	744	CCCCGGCGAGCTCACGGGGGCTTTGTAAAGCCATGAGAGCTGAGCGGCTCCGACCCGGG	803
OY	993	CGAGACGATTGGGAAGAAGACCGCTTCAGAGCGCTTGGAGCCGGCAACCGTTATCGAAT	1052
Db	804	CGAGACGATTGGGAAGAAGACCGCTTCAGAGCGCTTGGAGCCGGCAACCGTTATCGAAT	863
OY	1053	CCTTTCGGGACCCGGGTATTTGGGGGCTTCGGCGCTGAGGTGATCTTACAGAAACCCGA	1112
Db	864	CCTTTCGGGACCCGGGTATTTGGGGGCTTCGGCGCTGAGGTGATCTTACAGAAACCCGA	923
OY	1113	CGGCAAGCCGCAACGAAGAATTGAGGGTTACCCCATTCAGCGGAGCCGATCAGCTCG	1172
Db	924	CGGCAAGCCGCAACGAAGAATTGAGGGTTACCCCATTCAGCGGAGCCGATCAGCTCG	983
OY	1173	GCCGCTGAGCTTGAATTTCCGAGACCGATCATCGAGCCCGCTGAGTGGTATGACTTCAGC	1232
Db	984	GCCGCTGAGCTTGAATTTCCGAGACCGATCATCGAGCCCGCTGAGTGGTATGACTTCAGC	1043
OY	1233	GTTGGTTGGAACGGCAAGGGGGCGCGCAAGGGGCTTTCCCGGGGGCAAGCCATTCTCCG	1293
Db	1044	GTTGGTTGGAACGGCAAGGGGGCGCGCAAGGGGCTTTCCCGGGGGCAAGCCATTCTCCG	1103
OY	1293	CATGACAAAGCTGTACTGCGAGTGTGGCGCGTCATGACTTCGAAACGGGGGAAGAATC	1352
Db	1104	CATGACAAAGCTGTACTGCGAGTGTGGCGCGTCATGACTTCGAAACGGGGGAAGAATC	1166
OY	1353	GATCAAGACACTTACCGGCTGCCCTGCCCGAAGGTGTGAGACCCCTCCGACCTGGGCA	1412
Db	1166	GATCAAGACACTTACCGGCTGCCCTGCCCGAAGGTGTGAGACCCCTCCGACCTGGGCA	1222
OY	1413	GCAGAGGCAACGTTGCAACGTCAGCAATGCGGCACTTCGACAAAGTCTGTTGCGGAACGAT	1472
Db	1224	GCAGAGAGGCAACGTTGCAACGTCAGCAATGCGGCACTTCGACAAAGTCTGTTGCGGAACGAT	1283
OY	1473	CTTCAACAAGTCTAGGCAACGGCCGAAGGGCGCAAGAAAGACGTTGGCGCTCTTGGGAAGC	1533
Db	1284	CTTCAACAAGTCTAGGCAACGGCCGAAGGGCGCAAGAAAGACGTTGGCGCTCTTGGGAAGC	1343
OY	1533	CGCCGACGCTTTCGGCAAGCTCATCTGAGGGGCTTCGGAAGAAGGGGCAACGCAACT	1593
Db	1344	CGCCGACGCTTTCGGCAAGCTCATCTGAGGGGCTTCGGAAGAAGGGGCAACGCAACT	1403
OY	1593	TGTTTCGGAGCGCGCAAGCGCCCTGAAGCGCCCTTGAAGAAGCTGTACGAAGACCGCGGCG	1653
Db	1404	TGTTTCGGAGCGCGCAAGCGCCCTGAAGCGCCCTTGAAGAAGCTGTACGAAGACCGCGGCG	1463
OY	1653	AGGAGCTTACAGACGAGACCGCTTGGCAGGAAGACCTTCGGAAGCAAGACGACGCGCTGAC	1712
Db	1464	AGGCGCTTACAGACGAGACCGCTTGGCAGGAAGACCTTCGGAAGCAAGACGACGCGCTGAC	1523

QY 1713 GCTCCGACAGAGGGGCGGAAGAGCGCTTGGCGAAGCTTGAAGCGCCGGAAGCCCGAA 1772  
 Db 1524 GCTCCGACAGAGGGGCGGAAGAGCGCTTGGCGAAGCTTGAAGCGCCGGAAGCCCGAA 1583  
 QY 1773 GTTGGCCCTTGACCAATGTTCCCGAAGAGCGCGACCGCTACCCGACCGCCCTAAGTC 1832  
 Db 1584 GCTTCCCTTGACCAATGTTCCCGAAGAGCGCGACCGCTACCCGACCGCCCTAAGTC 1643  
 QY 1833 GTGTGGGGGGCGCGCTCAGTAGAGACAGACGCGGTGTTGCTCGGCTCTTCTAGACAA 1892  
 Db 1644 GTGTGGGGGGCGCGCTCAGTAGAGACAGACGCGGTGTTGCTCGGCTCTTCTAGACAA 1703  
 QY 1893 GATGCTTCTACAGAACTGACTACGGGAGGGGCGAGGAAAGCCCATCGAAGACGGCGC 1952  
 Db 1704 GATGCTTCTACAGAACTGACTACGGGAGGGGCGAGGAAAGCCCATCGAAGACGGCGC 1763  
 QY 1953 TTGATCAGCTGGGCGAAGCGCGACGACGACGACGACGACGACGACGACGACGACGAC 2012  
 Db 1764 TTGATCAGCTGGGCGAAGCGCGACGACGACGACGACGACGACGACGACGACGACGAC 1822  
 QY 2013 GGAAGACGTAG 2023  
 Db 1823 GGAAGACGTAG 1833

## RESULT 10

AAD19829  
 ID AAD19829 standard; DNA; 912 BP.

AC AAD19829;  
 DT 18-DEC-2001 (first entry)  
 DE Green fluorescent protein-1, syngFP1 gene with ST-LS1 intron sequence.  
 KW Cestrum yellow leaf curling virus; CMYLCV; transcription;  
 XX transgenic plant; green fluorescent protein 1; GFP1; ds.  
 OS Solanum tuberosum.  
 XX Unidentified.  
 PH Key Location/Qualifiers  
 FT Intron 278..465  
 FT /tag- a  
 PN /note- "Solanum tuberosum ST-LS1 intron"  
 WO200173087-A1.  
 XX 04-OCT-2001.  
 PD 26-MAR-2001; 2001WO-EP03408.  
 PF 27-MAR-2000; 2000GB-0007427.  
 PR 28-APR-2000; 2000GB-0010486.  
 PR 26-JAN-2001; 2001EP-0101802.  
 XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 PA Hohn T, Stavolone L, De Haan PT, Ligon HT, Kononova M;  
 PI WPI; 2001-616524/71.  
 DR 2001-616524/71.  
 XX Novel DNA sequence obtained from genome of Cestrum yellow leaf curling  
 PT virus for conferring constitutive expression of an associated desired  
 PT polynucleotide -  
 XX Example 19; Page 66; 100pp; English.  
 PS The invention relates to Cestrum yellow leaf curling virus (CMYLCV) novel  
 CC DNA sequences which functions as transcription promoters of an associated  
 CC polynucleotide sequence. These CMYLCV DNA molecules confers constitutive  
 CC expression of associated polynucleotide sequences. The invention also  
 CC relates to recombinant DNA sequences containing promoter sequences which

CC are used for creating transgenic plants expressing DNA of interest at all  
 CC times and in most tissues and organs. The present DNA sequence is green  
 CC fluorescent protein-1, syngFP1 gene with ST-LS1 intron sequence. Plant  
 CC optimised GFP reporter gene is used in the construction of plant  
 CC transformation vectors.  
 CC

Sequence 912 BP; 245 A; 245 C; 240 G; 182 T; 0 other;

## Query Match

9.8%; Score 198.6; DB 22; Length 912;

Best Local Similarity 95.8%; Pred. No. 4..8e-35;

Matches 204; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 21 CGGGGTGATACGTAAGTTCTGCTTCTTACCTTGAATATATATATATATATATATATAT 80  
 Db 265 CGGAGGCGTACGTAAGTTCTGCTTCTTACCTTGAATATATATATATATATATATATAT 324  
 QY 81 TTGATCTAAT 140  
 Db 325 TTGATCTAAT 384  
 QY 141 ATTGCTTTCTGATGTTATATAGTGTATATATATATATATATATATATATATATATAT 200  
 Db 385 ATTGCTTTCTGATGTTATATAGTGTATATATATATATATATATATATATATATATAT 444  
 QY 201 CCAAAATTGTGATGTGACGTACGCGGGTGC 233  
 Db 445 CCAAAATTGTGATGTGACGTACGCGGGTGC 477

## RESULT 11

AAD19834  
 ID AAD19834 standard; DNA; 1577 BP.

AC AAD19834;  
 DT 18-DEC-2001 (first entry)  
 DE Promoter-reporter cassette #3 to construct plant transformation vector.  
 KW Cestrum yellow leaf curling virus; CMYLCV; transcription;  
 XX transgenic plant; nopalline synthase; NOS; green fluorescent protein-1;  
 KW GFP1; ds.  
 XX Chimeric - Cestrum yellow leaf curling virus.  
 OS Chimeric - Agrobacterium tumefaciens.  
 OS Chimeric - Unidentified.  
 PH Key Location/Qualifiers  
 FT Promoter 1..354  
 FT /tag- a  
 FT /note- "CMYLCV CmpC promoter fragment"  
 FT misc\_feature 380..1292  
 FT /tag- b  
 FT /note- "Corresponds to SyngFP1 gene"  
 FT terminator 1304..1577  
 FT /tag- c  
 FT /note- "Nopalline synthase (NOS) terminator"  
 WO200173087-A1.  
 XX 04-OCT-2001.  
 PD 26-MAR-2001; 2001WO-EP03408.  
 PF 27-MAR-2000; 2000GB-0007427.  
 PR 28-APR-2000; 2000GB-0010486.  
 PR 26-JAN-2001; 2001EP-0101802.  
 XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 PA Hohn T, Stavolone L, De Haan PT, Ligon HT, Kononova M;  
 PI WPI; 2001-616524/71.  
 DR 2001-616524/71.



ID	AAJ19830 standard; DNA; 2001 BP.
XX	
AC	AAJ19830;
XX	
DT	18-DEC-2001 (first entry)
XX	
DE	Beta-glucuronidase (GUS) reporter gene with ST-LSI intron sequence.
XX	
KW	Cestrum yellow leaf curling virus; CmLYCV; transcripction;
KM	transgenic plant; beta-glucuronidase; GUS; ds.
XX	
OS	Solanum tuberosum.
XX	
XX	Unidentified.
FH	
FT	Key Location/Qualifiers
FT	Intron 385..576
FT	/tag= a
XX	/note= "Solanum tuberosum ST-LSI intron"
PN	WO200173087-A1.
XX	
PD	04-OCT-2001.
XX	
PF	26-MAR-2001; 2001WO-EP03408.
XX	
PR	27-MAR-2000; 2000GB-0007427.
PR	28-APR-2000; 2000GB-0010486.
PR	26-JAN-2001; 2001EP-0101802.
XX	
PA	(SYGN ) SYNGENTA PARTICIPATIONS AG.
XX	
PI	Hohn T, Stavolone L, De Haan PT, Ligou HT, Kononova M,
DR	WPI: 2001-616524/71.
XX	
PT	Novel DNA sequence obtained from genome of Cestrum yellow leaf curling
XX	virus for identifying constitutive expression of an associated desired
PS	polynucleotide
XX	
XX	Example 19; Page 67-68; 100pp; English.
CC	
CC	The invention relates to Cestrum yellow leaf curling virus (CmLYCV) novel
CC	DNA sequences which functions as transcription promoters of an associated
CC	polynucleotide sequence. These CmLYCV DNA molecules confers constitutive
CC	expression of associated polynucleotide sequences. The invention also
CC	relates to recombinant DNA sequences containing promoter sequences which
CC	are used for creating transgenic plants expressing DNA of interest at all
CC	times and in most tissues and organs. The present DNA sequence is
CC	beta-glucuronidase (GUS) reporter gene with ST-LSI intron sequence.
CC	The GUS reporter gene with intron) gene is used in the construction
CC	of plant transformation vectors.
XX	
XX	Sequence 2001 BP; 532 A; 451 C; 535 G; 483 T; 0 other:
XX	
QY	
DB	Query Match 9.6%; Score 195; DB 22; Length 2001;
QY	Best Local Similarity 100.0%; Pred. No. 4.2e-34;
DB	Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	
DB	30 TACGAAGTTTCGGTTCACCTTGATATATATTATAATTAATCACTAATTAGACTA 89
DB	
DB	383 TACGAAGTTTCGTCTTACCTTGATATATATTAATTAATCACTAATTAGACTA 442
QY	
DB	90 ATATAAATTTCCAATATTTTTTCCAATAAAGAAGATAGTATATGCAATTCCTTT 149
DB	
DB	443 ATATAAATTTCCAATATTTTTTCCAATAAAGAAGATAGTATATGCAATTCCTTT 502
QY	
DB	150 CTGAGTTTAAAGTGTATATTTTAATTATAAACCTTTCTAATATATGACCANAATT 209
DB	
DB	503 CTGAGTTTAAAGTGTATATTTTAATTATAAACCTTTCTAATATATGACCANAATT 562
QY	
DB	210 GTTGATGTGCAGTA 224
DB	
DB	563 GTTGATGTGCAGTA 577





GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 17:39:39 ; Search time 3307.5 Seconds

(without alignments)  
9944.984 Million cell updates/sec

Title: US-09-940-550a-10

Perfect score: 2031  
Sequence: 1 atggcacaagggtgtgtgac.....cggagacgtagcgcgtag 2031

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estda:\*  
2: em\_esthum:\*  
3: em\_estln:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inu:\*  
20: em\_gss\_pla:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	125.8	6.2	675	13	BM109636 EST557172
c 2	125.8	6.2	804	14	BQ512568 EST619983
3	125.8	6.2	847	14	BQ512569 EST619984
4	60.8	3.0	621	17	CNS04POX AL301578 Tetradon
c 5	58	2.9	928	17	CNS00DKY AL071865 Drosophila
c 6	58	2.9	964	17	CNS017VX AL108567 Drosophila

7	57.4	2.8	752	17	BH585665	BH585665	BOGSL20TF
c 8	57.4	2.8	986	17	CNS01707	AL107447	Drosophila
9	56.8	2.8	1101	17	CNS0175V	AL108460	Drosophila
c 10	56	2.8	780	17	CNS00774	AL083326	Arabidops
11	56	2.8	1201	17	CNS01038	AL098462	Drosophila
c 12	55.8	2.7	1101	17	CNS002AL	AL097431	Drosophila
c 13	55.6	2.7	1190	17	CNS020N7	AL206908	Tetradon
c 14	55.4	2.7	1101	17	CNS003BD	AL064091	Drosophila
c 15	55.2	2.7	925	17	CNS0091F	AL053013	Drosophila
16	55.2	2.7	975	17	BH179465	BH179465	014_P-10-
17	55	2.7	644	17	AQ961004	AQ961004	LERF158TF
c 18	55	2.7	720	17	CNS00X28	AL095522	Arabidops
c 19	54.8	2.7	859	17	BF272342	BF272342	GA_EB001
c 20	54.2	2.7	1101	17	CNS00EVL	AL069706	Drosophila
c 21	54.2	2.7	1101	17	CNS00EVL	AL069706	Drosophila
c 22	54	2.7	920	17	CNS01UOL	AL168150	Tetradon
c 23	53.8	2.6	602	17	B28108	B28108	T2G12TRD 7A
c 24	53.8	2.6	1101	17	CNS00CYH	AL060100	Drosophila
c 25	53.8	2.6	1101	17	CNS002DB	AL062360	Drosophila
c 26	53.4	2.6	454	17	B96681	B96681	T3001TR TAM
c 27	53.4	2.6	748	17	AQ286687	AQ286687	RFC111-81
c 28	53.4	2.6	876	17	CNS00961	AL053529	Drosophila
c 29	53.2	2.6	272	14	BQ596436	BQ596436	PTEStoab3
c 30	53.2	2.6	423	17	CNS00S62	AL087992	Arabidops
c 31	53.2	2.6	807	17	BH438258	BH438258	BOGL267TF
c 32	53.2	2.6	812	17	BH178455	BH178455	011_J-02-
c 33	53.2	2.6	812	17	CNS07KRM	AL615412	T3 end of
c 34	53.2	2.6	1101	17	CNS016LI	AL106896	Drosophila
c 35	53	2.6	364	17	AQ583787	AQ583787	RPCI-11-4
c 36	53	2.6	1101	17	CNS00DP7	AL075293	Drosophila
c 37	52.6	2.6	1101	17	CNS00EPO	AL069493	Drosophila
c 38	52.4	2.6	673	17	CNS00DWL	AL416395	T7 end of
c 39	52.4	2.6	905	17	CNS00KHX	AL077798	Drosophila
c 40	52.2	2.6	863	12	BF264892	BF264892	HY_CBA001
c 41	52.2	2.6	1101	17	CNS00EOL	AL069526	Drosophila
c 42	52	2.6	996	17	CNS00F0H	AL071063	Drosophila
c 43	52	2.6	1043	17	CNS0145P	AL103735	Drosophila
c 44	52	2.6	1292	13	BM463105	BM463105	ACBNCOURT
c 45	51.8	2.6	432	17	AQ293633	AQ293633	HS_2244_B

## ALIGNMENTS

RESULT 1  
BM109636/c 675 bp mRNA linear EST 26-NOV-2001  
LOCUS EST557172 potato roots Solanum tuberosum cDNA clone cPRO4H6 5' end,  
DEFINITION mRNA sequence.  
ACCESSION BM109636  
VERSION BM109636.1 GI:17070714  
KEYWORDS EST.  
SOURCE potato.  
ORGANISM Solanum tuberosum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
REFERENCE 1 (bases 1 to 675)  
van der Hoeven,R., Sun,H., Karamycheva,S.A., Tsai,J., Van Aken,S.,  
Ullrichback,T., Chienlung,A., Bougri,O., Buell,C.R., Ronning,C.,  
Tanksley,S. and Baker,B.  
Generation of ESTs from potato roots  
Unpublished (2001)  
Contact: Research Genetics, Libraries Division  
Tel: 1-800-711-6195  
Email: cdna@resgen.com  
For clone info: please contact Research Genetics, Libraries  
Division tel 1-800-711-6195, email cdna@resgen.com  
Seq primer: T3.  
Location/Qualifiers  
1..675  
/organism="Solanum tuberosum"  
/cultivar="Kennebec"

Query Match	6.28;	Score 125.8;	DB 13;	Length 675;
Best Local Similarity	-89.98;	Pred. No. 3.3e-18;		
Matches 167;	Conservative	0;	Mismatches 27;	Indels 5;
				Gaps 2

Accession	Result 2	LOCUS	DEFINITION
B0512568/c	B0512568/c	804 bp	linear
BQ512568	EST19983	Generation of a set of potato cDNA clones for microarray analyses	EST 10-JUN-2002
5' end	5' end	RNA sequence.	clone STM222

TITLE	Generation of a set of potato cDNA clones for microarray analyses
JOURNAL	Unpublished (2002)
COMMENT	Other_ESTs: EST619984

FEATURES  
source  
The Institute for Genomic Research  
9712 Medical Center Dr, Rockville, MD 20850, USA  
Email: potatob@tigr.org  
This clone is available through the Research Genetics, contact the  
Research Genetics for further information 1-800-711-6195 or  
cdna@rsngen.com  
Seq primer: 13.  
Location/Qualifiers  
1..804

```
Location/Qualifiers
1..804
/organism="Solanum tuberosum"
/cultivar="Kennebec Or Binjle"
/db_xref="taxon:4113"
/clone="STM422"
/clone_lib="generation of a set of potato cDNA clones for microarray analyses mixed potato tissues"
```

Query Match	6.2%	Score 125.8	DB 14	Length 804
Best Local Similarity	83.9%	Pred. No. 3.4e-18		
Matches 167; Conservative	0	Mismatches 27	Indels 5	Gaps 2

Accession	LOCUS	DEFINITION
RESULTS 3	BO512569	BO512569 847 bp mRNA linear EST 10-JUN-2002
		EST619984 Generation of a set of potato cDNA clones for microarray
		analyses mixed potato tissues Solanum tuberosum cDNA clone STMH22
		3' end, mRNA sequence.

ACCESSION VERSION KEYWORDS	SOURCE ORGANISM	TITLE JOURNAL COMMENT
BO512569 BO512569.1 EST.	potato. Solanum tuberosum	Generation of a set of potato cDNA clones for microarray analyses unpublished (2002) Other_ESTs: EST619983
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asterales; euasterids I; Solanales; Solanaceae; Solanum. 1 (bases 1 to 847)	
REFERENCE AUTHORS	Buelli, C. R., Hart, A., Baker, B., Tanksley, S., Fry, W., Smart, C., Restrepo, S., Griffiths, H., van der Hoeven, R., Tsai, J. and Karymchewa, S. A.	

FEATURES  
source  
The Institute for Genomic Research  
9712 Medical Center Dr, Rockville, MD 20850, USA  
Email: potatob@tigr.org  
This clone is available through the Research Genetics, contact the  
Research Genetics for further information 1-800-711-6195 or  
cdna@resgen.com  
Seq primer: 17'  
Location/Qualifiers  
1..847

```

location/Qualifiers
1. 847
/organism="Solanum tuberosum"
/cultivar="Kennebec or Bintje"
/db_xref="taxon:4113"
/clone="STMW22"
/clone_lib="generation of a set of potato cDNA clones for
microarray analyses mixed potato tissues"
/tissue_type="mixed tissues"
/lab_host="SOLR"
/note="Vector: pluscript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Combination of untreated and Phytophthora
infestans-treated libraries of roots, leaves, leaflets

```











[illegible][illegible]

VERSION AL053013.1 GI:4934461  
KEYWORDS GSS.  
SOURCE Drosophila melanogaster.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephyroidae; Drosophilidae; Drosophila.  
1 (bases 1 to 925)  
REFERENCE Genoscope.  
AUTHORS Direct Submission  
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
COMMENT Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see [http://www.fruitfly.org/TheBDGP/Drosophila  
melanogaster BAC library was prepared by Kazutoyo Osoegawa and  
Aaron Mammoser in Pieter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPC1-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain v2: cn bw sp, the same strain used for the BDGP's  
PL and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at \[http://bacpac.med.buffalo.edu/drosophila\\\_bac.htm\]\(http://bacpac.med.buffalo.edu/drosophila\_bac.htm\).  
Location/Qualifiers  
1. 925  
/organism="Drosophila melanogaster"  
/db\\_xref="taxon:7227"  
/clone="BACRPC198"  
/clone\\_lib="RPC1-98"  
/note="end : TERT3"  
BASE COUNT 120 a 61 c 61 g 172 t 511 others  
ORIGIN  
Query Match 2.7%; Score 55.2; DB 17; Length 925;  
Best Local Similarity 13.6%; Pred. No. 0.052;  
Matches 51; Conservative 171; Mismatches 154; Indels 0; Gaps 0;  
QY 1516 GCGCTTCTGTGGAGACGCCGACGCTTGGCAAGCTCAGTGAAGCGCTGGAAGAC 1575  
DB 924 SBSCSCSCSCSSBSCSSSMSTSSMSBSCSSBSSSTSSMSBSSSSGSSSS 865  
QY 1576 GCGGAACGGCGAAGCTTGTGCGAGCGCGGACGCCCTGACGCCCTTGAAGAGCTG 1635  
DB 864 SGTSSACVNCNASSSCGCCGACBACBMCSSSSSCGASARGVAVRAGAGKRGCGSG 805  
QY 1636 TACGAAGACCGCGGCGGACGAGTACGACGACCGCTTGGCAGAGACGACTTCCGAG 1695  
DB 804 GASASHSSSSACBSSSSSCASCMWSASSSSASRSRGAGGSGGASSSRSSSSSA 745  
QY 1696 CACAGAGCGAGCTGACGCTCCGAGAGCAAGAGGCGGAGAGCGGCTTCCCACTTGA 1755  
DB 744 SAGSYVSSASSSSSSSCSSSVASSMSCSBSSSSASASSSSSSSSASACSCCCT 685  
QY 1756 GCGCGCGAAGCCCGAAGTCCCTTGACCAATGTTCCCGAAGACCGCAGCTGAC 1815  
DB 684 SWSGSCSTISASMSARSSSSSSSSSSMSASASSSSASASSSSSSSSSSSSGSA 625  
QY 1816 CCGACCGCGCTTAAGTGTGAGTGGGGGCGCGGCTAGTAGACAGACGCGCTGTC 1875  
DB 624 MSSGGGSGSVSSMSMSSSVSSGGRSSGSGGGGAGGSSGSSGSSGSSGSSVCS 565  
QY 1876 GGGCTCTTCTGAGACA 1891  
DB 564 GCMCRSCSSAAAAA 549](http://www.fruitfly.org/TheBDGP/Drosophila%20melanogaster%20BAC%20library%20prepared%20by%20Kazutoyo%20Osoegawa%20and%20Aaron%20Mammoser%20in%20Pieter%20de%20Jong's%20laboratory%20in%20the%20Department%20of%20Cancer%20Genetics%20at%20the%20Roswell%20Park%20Cancer%20Institute%20in%20Buffalo,%20NY.%20The%20library%20is%20named%20RPC1-98%20and%20was%20constructed%20by%20partial%20EcoRI%20digestion%20of%20Drosophila%20DNA%20provided%20by%20the%20BDGP%20from%20the%20isogenic%20strain%20v2%20cn%20bw%20sp,%20the%20same%20strain%20used%20for%20the%20BDGP's%20PL%20and%20EST%20libraries.%20A%20more%20detailed%20description%20of%20the%20library%20and%20how%20to%20order%20individual%20BAC%20clones,%20the%20entire%20library,%20or%20filters%20for%20hybridization%20from%20the%20BACPAC%20Resource%20Center%20can%20be%20found%20at%20http://bacpac.med.buffalo.edu/drosophila_bac.htm)

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 17:41:53 ; Search time 107 Seconds

(without alignments)  
5821.126 Million cell updates/sec

Title: US-09-940-550A-10

Perfect score: 2031  
Sequence: 1 atggcacaaggggtgtgtac.....cgaagacgtacgagcgtag 2031

Scoring table: IDENTITY-NUC

Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database:

Issued\_Patents\_NA: \*  
1: /cgn2\_6/ptodata/1/lna/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/1/lna/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/1/lna/6A\_COMB.seq: \*  
4: /cgn2\_6/ptodata/1/lna/6B\_COMB.seq: \*  
5: /cgn2\_6/ptodata/1/lna/PTCUTS\_COMB.seq: \*  
6: /cgn2\_6/ptodata/1/lna/backfillseq1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1796.6	88.5	2091	6	5190871-3
2	1793.4	88.3	3401	6	5190871-1
3	195	9.6	2633	1	US-08-452-267-2
4	195	9.6	2633	1	US-09-123-644-2
5	195	9.6	2633	1	US-08-452-267-3
6	195	9.6	5534	3	US-09-123-644-3
7	194	9.6	1701	4	US-09-080-625-1
8	194	9.6	1701	4	US-09-080-625-2
9	194	9.6	3336	4	US-09-080-625-2
10	194	9.6	3336	4	US-09-080-625-2
11	194	9.6	3877	4	US-09-080-625-4
12	194	9.6	3877	4	US-09-080-625-4
13	190.2	9.4	5560	3	US-08-695-782-4
14	190	9.4	1034	4	US-08-817-188-5
15	51.2	2.5	19124	2	US-09-367-293-2
16	51.2	2.5	19124	2	US-08-487-826B-13
17	51	2.5	1926	4	US-09-249-585A-4
18	46.8	2.3	1931	2	US-09-130-114-2
19	46.6	2.2	1931	2	US-08-487-826B-13
20	45	2.2	722	4	US-08-998-416-780
21	45	2.2	727	4	US-08-998-416-780
22	44.8	2.2	8700	2	US-08-466-961A-16
23	44.8	2.2	8700	2	US-08-645-193B-18
24	44.4	2.2	5923	4	US-09-064-922-3
25	44.4	2.2	2960	3	US-08-913-842-3
26	44	2.1	1850	3	US-08-617-860B-32
27	43.4	2.1	1850	3	US-08-617-860B-32

28	43.4	2.1	4098	2	US-08-605-106-4	Sequence 4, Appli
29	43	2.1	5852	1	US-07-867-106-2	Sequence 2, Appli
30	42.6	2.1	6265	1	US-09-129-112-3	Sequence 2, Appli
31	42.2	2.1	1431	4	US-09-316-083-2	Sequence 2, Appli
32	42	2.1	615	4	US-08-998-416-186	Sequence 18, App
33	41.8	2.1	7218	1	US-08-232-463-14	Sequence 14, Appl
34	41.4	2.0	7815	1	US-09-102-528-28	Sequence 28, Appl
35	41.4	2.0	20235	1	US-07-642-734C-3	Sequence 3, Appli
36	41.4	2.0	20235	3	US-08-439-009A-3	Sequence 3, Appli
37	41.4	2.0	168575	4	US-09-426-290-1	Sequence 1, Appli
38	41.4	2.0	4403765	4	US-09-103-840A-2	Sequence 2, Appli
39	41	2.0	2960	3	US-08-913-842-3	Sequence 36, Appli
40	40.8	2.0	3618	1	US-07-872-678A-36	Sequence 3, Appli
41	40.8	2.0	4411529	4	US-09-103-840A-1	Sequence 1, Appli
42	40.6	2.0	1150	1	US-08-161-406-1	Sequence 10, Appl
43	40.4	2.0	1600	4	US-09-434-288-10	Sequence 4, Appli
44	40.2	2.0	3196	2	US-09-096-982-4	Sequence 4, Appli
45	40.2	2.0	3196	2	US-08-653-650A-4	Sequence 4, Appli

## ALIGNMENTS

RESULT 1	5190871-3	Patent No. 5190871-3	APPLICANT: COX, KAREN L.; KUHSTOSS, STUART A.; RAO, R. NAGARAJA; RICHARDSON, MARK A.; SCHONER, BRIGITTE E.; SENO, EUGENE T.
FUNCTION OF INVENTION: USE OF THE SITE-SPECIFIC INTEGRATING			
NUMBER OF SEQUENCES: 3			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/07/364,959			
FILING DATE: 12-JUN-1989			
SEQ ID NO: 3			
LENGTH: 2091			
5190871-3			
Query Match	88.5%	Score 1796.6;	DB 6; Length 2091;
Best Local Similarity	99.2%	Pred. No. 0;	
Matches 1805; Conservative	0;	Mismatches 14; Indels 0; Gaps 0;	
213 GATGTCAGTACGCGGGGCTTACAGCCGTCAGTCGCGAGCGGAGAAATGACAGTC 272			
254 GGTGACACAGTACGCGGGGCTTACAGCCGTCAGTCGCGAGCGGAGAAATGACAGTC 313			
273 ACCAAGCCGACGACAGCGGTAGCGCCCAAGCAAGAGCGCGGACCTTACGCGCA 332			
314 ACCAAGCCGACGACAGCGGTAGCGCCCAAGCAAGAGCGCGGACCTTACGCGCA 373			
333 ACTCGAGCGGACGCGGGGCTTACAGTCGCGGACCTTACAGGAGCGCGGAC 392			
374 AGTCGAGCGGACGCGGGGCTTACAGTCGCGGACCTTACAGGAGCGCGGAC 433			
393 GTCGCGGTCGCGGACGCGCGGAGCGCCGAGTTCGAACGATTCGAAGATTCGCGCGC 452			
434 GTCGCGGTCGCGGACGCGCGGAGCGCCGAGTTCGAACGATTCGAAGATTCGCGCGC 493			
453 CGGCGGCTCAACATGATTCGTCATGACGTCGCGCTTCGCGGCTCAAGGTCAT 512			
494 CGGCGGCTCAACATGATTCGTCATGACGTCGCGCTTCGCGGCTCAAGGTCAT 553			
513 GCAAGCGATTCGATTCGTCGGAATTCGCGGCTTCGCGGCTTCGAGATTCGCGCTCA 572			
554 GCAAGCGATTCGATTCGTCGGAATTCGCGGCTTCGCGGCTTCGAGATTCGCGCTCA 613			
573 GGAAGCGCTTCGCGGACGAGAAAGTCATGACGTCGATTCGCGGATTCGCGGCTCA 632			
614 GGAAGCGCTTCGCGGACGAGAAAGTCATGACGTCGATTCGCGGATTCGCGGCTCA 673			
633 CGCGTCGCAAGAAATTCGTCGTAAGTCGCGGAGAAATTCGCAACGAAACCTTCA 692			
674 CGCGTCGCAAGAAATTCGTCGTAAGTCGCGGAGAAATTCGCAACGAAACCTTCA 733			

```

QY 693 GCGCAATTGGGCGGTACGTGGGCGGGAAGGCGCTTACGGCTTCAGCTTGTTCGGA 752
DB 734 GGGCAATTTGGGCGGTAGTGGGCGGGAAGGCGCTTACGGCTTGTTCGGA 793
QY 753 GACGAAGGATCACGCGGCAACGCGGATGTCATGTCATCAACAGTTAGGCGCA 812
DB 794 GACGAAGGATCACGCGGCAACGCGGATGTCATGTCATCAACAGTTAGGCGCA 853
QY 813 CTCGACCACTCCCTTACCGGACCTTCGAGTTGAGCCCGGAGTAATCCGTTGGTGG 872
DB 854 CTCGACCACTCCCTTACCGGACCTTCGAGTTGAGCCCGGAGTAATCCGTTGGTGG 913
QY 873 GCGTAGATCAAGACGCAACACCTTCCTTCAACCCGCGGAGTCAGACCCGCAATTA 932
DB 914 GCGTAGATCAAGACGCAACACCTTCCTTCAACCCGCGGAGTCAGACCCGCAATTA 973
QY 933 CCGCGGAGCATCAAGGCGGCTTTGTATAGGCGATGAGCGCTGAGCGGCTCCGACCGGG 992
DB 974 CCGCGGAGCATCAAGGCGGCTTTGTATAGGCGATGAGCGCTGAGCGGCTCCGACCGGG 1033
QY 993 CGAAGCATTTGGGAAGAACGCGCTTCAAGCGCTTGGACCGGCAACCGCTTATGCGAAT 1052
DB 1034 CGAAGCATTTGGGAAGAACGCGCTTCAAGCGCTTGGACCGGCAACCGCTTATGCGAAT 1093
QY 1053 CTTTCGGGACCCCGCTATTGCGGGCTTTCGCGGTGAGGATCTCAAGAGAAGCCGGA 1112
DB 1094 CTTTCGGGACCCCGCTATTGCGGGCTTTCGCGGTGAGGATCTCAAGAGAAGCCGGA 1153
QY 1113 CCGCAGCGGACGACCAAAATTTAGGGTTACCGCATTCAGCGGACCGCATCAGCTCCG 1172
DB 1154 CCGCAGCGGACGACCAAAATTTAGGGTTACCGCATTCAGCGGACCGCATCAGCTCCG 1213
QY 1173 GCGGCTGAGCTTGAATTTGGGACCGCATCAGCGGCTGAGGTTAGGCTTCAGGC 1232
DB 1214 GCGGCTGAGCTTGAATTTGGGACCGCATCAGCGGCTGAGGTTAGGCTTCAGGC 1273
QY 1233 GTGGTTGGACGGGCGGCGGCGGCAAGGCGCTTCCGGGGGCAAGCCATTTGTCGCG 1292
DB 1274 GTGGTTGGACGGGCGGCGGCGGCAAGGCGCTTCCGGGGGCAAGCCATTTGTCGCG 1333
QY 1293 CATGACAAAGCTTACTCGAGTGTGGGCGGCTCATGACTTCGAAGCGGCGGCAAGT 1352
DB 1334 CATGACAAAGCTTACTCGAGTGTGGGCGGCTCATGACTTCGAAGCGGCGGCAAGT 1393
QY 1353 GATCAAGGACTTTTACCGGCTCCGCTCGCGAAGGTTGTGACCCGCTCCGACCTGGGA 1412
DB 1394 GATCAAGGACTTTTACCGGCTCCGCTCGCGAAGGTTGTGACCCGCTCCGACCTGGGA 1453
QY 1413 GCACGAAGGACGTCAGACGTCAGGCGGACCTGCAAAATGTTGGTGGAGCGAT 1472
DB 1454 GCACGAAGGACGTCAGACGTCAGGCGGACCTGCAAAATGTTGGTGGAGCGAT 1513
QY 1473 CTTCAACAAATGTCAGGACCGGCAAGGCGGCAAGAGAGCTTGTGGGAAGC 1532
DB 1514 CTTCAACAAATGTCAGGACCGGCAAGGCGGCAAGAGAGCTTGTGGGAAGC 1573
QY 1533 CGCCCGACGCTTTCGGAAGTCACTAGGCGGCTTGAAGAGCGGCGAAGCGGCAACT 1592
DB 1574 CGCCCGACGCTTTCGGAAGTCACTAGGCGGCTTGAAGAGCGGCGAAGCGGCAACT 1633
QY 1593 TGTTCGGAGCGGCGGCGGCTTGAAGCGCTTGAAGAGCTTGAAGAGCGGCGGCG 1652
DB 1634 TGTTCGGAGCGGCGGCGGCGGCTTGAAGCGCTTGAAGAGCTTGAAGAGCGGCGGCG 1693
QY 1653 AGGAGCTTACGAAGCCGCTTGGAGAGAGCTTTCGGAAGCAAGCGAGCGCTGAC 1712
DB 1694 AGGAGCTTACGAAGCCGCTTGGAGAGAGCTTTCGGAAGCAAGCGAGCGCTGAC 1753
QY 1713 GCTTCGGAGCAAGGCGGGAAGAGCGGCTTTCGGAAGCAAGCGAGCGCTGAC 1772
DB 1754 GCTTCGGAGCAAGGCGGGAAGAGCGGCTTTCGGAAGCAAGCGAGCGCTGAC 1813

```

```

QY 1773 GTTGGCCCTTGAACCAATGTTCCCGGAAGACCGGACGCTGACCCGACCGCTTAGTGC 1832
DB 1814 GCTTCCCTTGAACCAATGTTCCCGGAAGACCGGACGCTGACCCGACCGCTTAGTGC 1873
QY 1833 GTGAGGGGGGCGCGCTCAGTAGAGACCAAGCGGCTTGTGCGGCTCTTCTAGCAAA 1892
DB 1874 GTGAGGGGGGCGCGCTCAGTAGAGACCAAGCGGCTTGTGCGGCTCTTCTAGCAAA 1933
QY 1893 GATCGTTTTCAGAAATGCTAGCTACGGGCGAGGGGCAAGGAGCCCATGAGAAAGCGGC 1952
DB 1934 GATCGTTTTCAGAAATGCTAGCTAGGGGCGAGGGGCAAGGAGCCCATGAGAAAGCGGC 1993
QY 1953 TTTCATACGTTGGGCGAAGCGCGGACCGGACCGAGCAAGCAAGAGCGCCAGAGCGGAC 2012
DB 1994 TTTCATACGTTGGGCGAAGCGCGGACCGGACCGAGCAAGCAAGAGCGCCAGAGCGGAC 2053
QY 2013 GGAAGACGTAAGCGGCGTAG 2031
DB 2054 GGAAGACGTAAGCGGCGTAG 2072

RESULT 2
5190871-1
; Patent NO. 5190871
; APPLICANT: COX, KAREN L.; KUHSTOSS, STUART A.; RAO, R. NAGARAJA
; RICHARDSON, MARK A.; SCHONER, BRIGITTE E.; SENO, EUGENE T.
; TITLE OF INVENTION: USE OF THE SITE-SPECIFIC INTEGRATING
; FUNCTION OF PHAGE C31
; NUMBER OF SEQUENCES: 3
; CURRENT APPLICATION NUMBER:
; APPLICATION NUMBER: US/07/364,959
; FILING DATE: 12-JUN-1989
; SEQ ID NO: 1
; LENGTH: 3401
5190871-1

Query Match: 88.3%; Score 1793.4; DB 6; Length 3401;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1803; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 213 GATGCGAGGTAACCGGCGGCTTTCAGACCGTCACTGCGGAGCGGAGTAATGACGTC 272
DB 532 GGTGACACGTAACCGGCGGCTTTCAGACCGTCACTGCGGAGCGGAGTAATGACGTC 272
QY 273 AGCAAGCCGACGACAGGCTAGCGGCAAGCAAGCAAGCAAGCGGCGGAGCTTACGCGCA 332
DB 592 AGCAAGCCGACGACAGGCTAGCGGCAAGCAAGCAAGCAAGCGGCGGAGCTTACGCGCA 332
QY 333 AGTGAAGCGGAGCGGCGGCGGCTTTCAGGCTTTCAGGCAATTCAGGCAAGCGGCGGAC 392
DB 652 AGTGAAGCGGAGCGGCGGCGGCTTTCAGGCTTTCAGGCAATTCAGGCAAGCGGCGGAC 392
QY 652 AGTGAAGCGGAGCGGCGGCGGCTTTCAGGCTTTCAGGCAATTCAGGCAAGCGGCGGAC 392
DB 692 AGTGAAGCGGAGCGGCGGCGGCTTTCAGGCTTTCAGGCAATTCAGGCAAGCGGCGGAC 392
QY 692 AGTGAAGCGGAGCGGCGGCGGCTTTCAGGCTTTCAGGCAATTCAGGCAAGCGGCGGAC 392
DB 712 GTGGGCGTTCGGAAGCGGCGGAGCGGCGGAGTTTCGAAGCAATTCGAAGCAATTCGCGC 771
QY 712 GTGGGCGTTCGGAAGCGGCGGAGCGGCGGAGTTTCGAAGCAATTCGAAGCAATTCGCGC 771
DB 771 GTGGGCGTTCGGAAGCGGCGGAGCGGCGGAGTTTCGAAGCAATTCGAAGCAATTCGCGC 771
QY 771 GTGGGCGTTCGGAAGCGGCGGAGCGGCGGAGTTTCGAAGCAATTCGAAGCAATTCGCGC 771
DB 832 GTGGGCGTTCGGAAGCGGCGGAGCGGCGGAGTTTCGAAGCAATTCGAAGCAATTCGCGC 831
QY 832 GTGGGCGTTCGGAAGCGGCGGAGCGGCGGAGTTTCGAAGCAATTCGAAGCAATTCGCGC 831
DB 892 GTGGGCGTTCGGAAGCGGCGGAGCGGCGGAGTTTCGAAGCAATTCGAAGCAATTCGCGC 891
QY 892 GTGGGCGTTCGGAAGCGGCGGAGCGGCGGAGTTTCGAAGCAATTCGAAGCAATTCGCGC 891
DB 951 GTGGGCGTTCGGAAGCGGCGGAGCGGCGGAGTTTCGAAGCAATTCGAAGCAATTCGCGC 951
QY 951 GTGGGCGTTCGGAAGCGGCGGAGCGGCGGAGTTTCGAAGCAATTCGAAGCAATTCGCGC 951
DB 991 GTGGGCGTTCGGAAGCGGCGGAGCGGCGGAGTTTCGAAGCAATTCGAAGCAATTCGCGC 991
QY 991 GTGGGCGTTCGGAAGCGGCGGAGCGGCGGAGTTTCGAAGCAATTCGAAGCAATTCGCGC 991
DB 1011 GTGGGCGTTCGGAAGCGGCGGAGCGGCGGAGTTTCGAAGCAATTCGAAGCAATTCGCGC 1011
QY 1011 GTGGGCGTTCGGAAGCGGCGGAGCGGCGGAGTTTCGAAGCAATTCGAAGCAATTCGCGC 1011

```



Db	1012	GC	GGAAATTGGGGGGGATAGCTCGGGCGGGAAGGGCCCTTACGGCTTCGAGCTTGTTGGCA	1071
Qy	753	GAC	CAGAGAGATCACGGCGCAACGGCGGAATGTCATATGCTCATACAAAGTTAGCCCA	812
Db	1072	GAC	CAGAGAGATCACGGCGCAACGGCGGAATGTCATATGCTCATACAAAGCTTGCCCA	1131
Qy	813	CTC	GACCACTCCCTTCACGGACCTTCGAGTTCGAGCCCGCAAGTAATCCGGTGGTGTG	872
Db	1132	CTC	GACCACTCCCTTCACGGACCTTCGAGTTCGAGCCCGCAAGTAATCCGGTGGTGTG	1191
Qy	873	GC	GTAGATCAAGCAGCAACAACCTTCCTTCAAGCCGGGAGTCAAGCCGCCATTC	932
Db	1192	GC	GTAGATCAAGCAGCAACAACCTTCCTTCAAGCCGGGAGTCAAGCCGCCATTC	1251
Qy	933	CCG	GGGAGCATCAGGGGGCTTGTAAAGGCATGAGCGCTGAGCGCTGCGAGCCGGG	992
Db	1252	CCG	GGGAGCATCAGGGGGCTTGTAAAGGCATGAGCGCTGAGCGCGCTGCGAGCCGGG	1311
Qy	993	CGA	GACGATTTGGGAAGAAGACCCCTTCAGAGCGCTTGGAACCCGGCAACCTGATGCAAT	1055
Db	1312	CGA	GACGATTTGGGAAGAAGACCCGCTTCAGAGCGCTTGGAACCCGGCAACCTGATGCAAT	1371
Qy	1053	CTT	TCGGGACCCCGCTATTTCGGGGCTTCGCCCTGAGGTGATCTACAGAGAGCCGA	1112
Db	1372	CTT	TCGGGACCCCGCTATTTCGGGGCTTCGCCCTGAGGTGATCTACAGAGAGCCGA	1431
Qy	1113	CGG	CAGCGCCGACCAAGAAATTGAGGGTTACCCGATTCAGCGCGAGCCGATCAGCTCG	1172
Db	1432	CGG	CAGCGCCGACCAAGAAATTGAGGGTTACCCGATTCAGCGCGAGCCGATCAGCTCG	1491
Qy	1173	GCC	GGTCAGCTTGATTTGGGAGCCGATCATGAGCCCGCTGAGTGGTATGACTTCAGGC	1232
Db	1492	GCC	GGTCAGCTTGATTTGGGAGCCGATCATGAGCCCGCTGAGTGGTATGACTTCAGGC	1551
Qy	1233	GTT	GTTGGACGGCGAGGGGGGGCGCAAGGGGCTTCCCGGGGGCAAGCATTCCTCGCG	1292
Db	1552	GTT	GTTGGACGGCGAGGGGGGGCGCAAGGGGCTTCCCGGGGGCAAGCATTCCTCGCG	1611
Qy	1293	CAT	GAGCAAGCTGTACTGCGCAAGTGTGGCGGCTCATGACTTGTGGAAGCGCGGAGAAATC	1352
Db	1612	CAT	GAGCAAGCTGTACTGCGCAAGTGTGGCGGCTCATGACTTGTGGAAGCGCGGAGAAATC	1671
Qy	1353	GAT	CAGGACTCTTACCGCTGCGGTGGCGGGAAGGTGTGTGACCGCTCCGACCTGGGCA	1412
Db	1672	GAT	CAGGACTCTTACCGCTGCGGTGGCGGGAAGGTGTGTGACCGCTCCGACCTGGGCA	1731
Qy	1413	GC	CAGAGAGGCACGTGAACGTCAGACATGGGGGACTCGACAAGTTGCTTGGCGAAGCGAT	1472
Db	1732	GC	CAGAGAGGCACGTGAACGTCAGACATGGGGGACTCGACAAGTTGCTTGGCGAAGCGAT	1791
Qy	1473	CTT	CAACAATCATAGGACCGCCGGAAGGCGACGAAGAAGACTTTGGCGCTTCTGTGGGAAGC	1532
Db	1792	CTT	CAACAATCATAGGACCGCCGGAAGGCGACGAAGAAGACTTTGGCGCTTCTGTGGGAAGC	1851
Qy	1533	CGC	CGCGAGCTTGGGCAAGCTCACTAGAGGGCTTGAGAAGACCGCGCAACGGGCGCAACT	1592
Db	1852	CGC	CGCGAGCTTGGGCAAGCTCACTAGAGGGCTTGAGAAGACCGCGCGCAACGGGCGCAACT	1911
Qy	1593	TGT	TTCGAGAGCGCCCAACCGCTTGAAAGCGCTTGAAGAAGCTGTACGAAGACCGCGCGG	1652
Db	1912	TGT	TTCGAGAGCGCGCAACCGCTTGAAAGCGCTTGAAGAAGCTGTACGAAGACCGCGCGG	1971
Qy	1653	AGA	GACTTACGAGCGACCCGTTGGCAGGAAGCACTTCCGGAAGCAACGAGCGCTGAC	1712
Db	1972	AGA	GACTTACGAGCGACCCGTTGGCAGGAAGCACTTCCGGAAGCAACGAGCGCTGAC	2031
Qy	1713	GCT	CCGAGCAAGGGCGGAAGAGCGGCTTGGCGCAACTTGAAGCGCGCGCAAGCCCGCA	1772
Db	2032	GCT	CCGAGCAAGGGCGGAAGAGCGGCTTGGCGCAACTTGAAGCGCGCGCAAGCCCGCA	2091
Qy	1773	GTT	GCCCCCTTGACCAATGTTCCCGGAAGACGCGACGCTGACCCGACCGGCTTAATGC	1832

Accession	Sequence	Length
Db 2092	GCTTCCCTTGACCAATGTTTCCCCGAAGACGCCGACGCTGAACCCGACGCTTAATG	2151
Oy 1833	GTGGTGGGGGGCGCCCGTCAGTAGACCAAGCCGCTGTTCGTGGGCTTTGTGTAGACA	1892
Db 2152	GTGGTGGGGGGCGCCCGTCAGTAGACCAAGCCGCTGTTCGTGGGCTTTGTGTAGACA	2211
Oy 1893	GATCGTGTCTCACAAGTCGACTACGAGCAGGGGGGACGGGACGCCCATGAGAACCGCG	1952
Db 2212	GATCGTGTCTCACAAGTCGACTACGAGCAGGGGGGACGGGACGCCCATGAGAACCGCG	2271
Oy 1953	TTTCGATCAGTGGGCGAAGCCGCGACCGACGACGACGAAGAGACGCCAGAGCGCAC	2012
Db 2272	TTTCGATCAGTGGGCGAAGCCGCGACCGACGACGACGAAGAGACGCCAGAGCGCAC	2331
Oy 2013	GGAAAGAGTACGGGCTAG 2031	
Db 2332	GGAAAGAGTACGGGCTAG 2350	

```

      RESULT 3
US-08-452-267-2
; Sequence 2, Application US/08452267
; Patent No. 5801027
; GENERAL INFORMATION:
; APPLICANT: Bennett, Malcolm
; APPLICANT: May, Sean
; APPLICANT: Ramsay, Nichola
; TITLE OF INVENTION: Control of Genes in Transgenic Plants
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Barnes & Thornburg
; STREET: 11 South Meridian
; CITY: Indianapolis
; STATE: IN
; COUNTRY: USA
; ZIP: 46204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452, 267
; FILING DATE: 26-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Breen, John P.
; REGISTRATION NUMBER: 38, 833
; REFERENCE/DOCKET NUMBER: 6653-25744
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 231-7745
; TELEFAX: (317) 231-7433
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2633 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; CLONE: pUMIGIT(insert)
; US-08-452-267-2

Query Match          9.6%; Score 195; DB 1; Length 2633;
Best Local Similarity 100.0%; Pred. No. 2.8e-36;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

    30 TACGTAAGTTTCGTCTTACCCTTGATTTAATAATTAAATTAATCAATTAATAGTAGTA 89
       |||.....|||.....|||.....|||.....|||.....|||.....|||.....
DB     693 TACGTAAGTTTCGTCTTACCCTTGATTTAATAATTAAATTAATCAATTAATAGTAGTA 752

```



```

APPLICANT : Kriz, Alan L.
TITLE OF INVENTION : METHODS AND COMPOSITIONS FOR TRANSGENE
IDENTIFICATION
NUMBER OF SEQUENCES : 19
CORRESPONDENCE ADDRESS :
ADDRESSEE : Arnold White & Durkee
STREET : P.O. Box 4433
CITY : Houston
STATE : TX
COUNTRY : USA
ZIP : 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/080,625
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, Robert E.
REGISTRATION NUMBER: P-42,628
REFERENCE/DOCKET NUMBER: DEKM:161
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1701 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-080-625-1

Query Match          9.6%; Score 194; DB 4; Length 1701;
Best Local Similarity 97.5%; Pred. No. 4e-38;
Matches 197; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      23 GGGGAGATACCTAAGTTCCTCTCTACCTTGATATATATATATATCATTAATT 82
Db      370 GAGCTGAAGTAGTTCTCTCTCTACCTTGATATATATATATATATATCAATT 429
QY      83 AGTAGTAAATATAATTTCAAATATTTTTTCCAAAATAAAGAATGTATATGCAAT 142
Db      430 AGTAGTAAATATAAATTTCCAATATTTTTTTCCAAAATAAAGAATGTATATGCAAT 489
QY      143 TGCCTTCTGTAGTTTAAAGTGTATATTTAATTATATACCTTCTATATATGACC 202
Db      490 TGCCTTCTGTAGTTTAAAGTGTATATTTTAAATTTATATACCTTCTATATATGACC 549
QY      203 AAAATTGTTGATGTGACAGTA 224
Db      550 AAAATTGTTGATGTGACAGTA 571

RESULT 8
US-09-695-782-1
Sequence 1, Application US/09695782
Patent No. 6433252
GENERAL INFORMATION:
APPLICANT: KRIZ, ALAN L.
APPLICANT: GRIFFOR, MATTHEW
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE IDENTIFICATION
FILE REFERENCE: DEKM:161USCI
CURRENT APPLICATION NUMBER: US/09/695,782
CURRENT FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: 09/080,625
PRIOR FILING DATE: 1998-05-18
NUMBER OF SEQ ID NOS: 19
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 1
```

```

QY      23 GGGTGGATCGTAAGTTCTGCTTCACCTTGATATATATAAATATCATTAATT 82
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1441 GAGCTGAAGGTAAAGTTTCGCCCTTCACCTTGATATATATAAATATCATTAATT 1500
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      83 AGTAGTAATATATATATTTCAATATTTTTTCAAATAAAGAAGTAGATATAGCAAT 142
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1501 AGTAGTAATATATATATTTCAATATTTTTTCAAATAAAGAAGTAGATATAGCAAT 1560
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      143 TGCCTTTCTGATGTTATTAAGTGTATATTTTAATTAACTTCTTAATATATGACC 202
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1561 TGCCTTTCTGATGTTATTAAGTGTATATTTTAATTAACTTCTTAATATATGACC 1620
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      203 AAAATTGTTGATGTCAGGTA 224
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1621 AAAATTGTTGATGTCAGGTA 1642
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 10
US-09-695-782-2          *
; Sequence 2, Application US/09695782
; Patent No. 6433252
; GENERAL INFORMATION:
; APPLICANT: KRIZ, ALAN L.
; APPLICANT: GRIFFOR, MATTHEW
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE IDENTIFICATION
; FILE REFERENCE: DEM:161USC1
; CURRENT APPLICATION NUMBER: US/09/695,782
; PRIOR FILING DATE: 2000-10-24
; PRIORITY APPLICATION NUMBER: 09/080,625
; PRIOR FILING DATE: 1998-05-18
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 3336
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Primer
US-09-695-782-2

Query Match              9.6%; Score 194; DB 4; Length 3336;
Best Local Similarity    97.5%; Pred. No. 5.4e-38;
Matches 197; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      23 GGGTGATACGTAACTTCTGCTTCACCTTGATATATATAAATATCATTAATT 82
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1441 GAGCTGAAGGTAAAGTTCTGCTTCACCTTGATATATATAAATATCATTAATT 1500
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      83 AGTAGTAATATATATATTTCAATATTTTTTCAAATAAAGAAGTAGATATAGCAAT 142
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1501 AGTAGTAATATATATATTTCAATATTTTTTCAAATAAAGAAGTAGATATAGCAAT 1560
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      143 TGCCTTTCTGATGTTATTAAGTGTATATTTTAATTAACTTCTTAATATATGACC 202
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1561 TGCCTTTCTGATGTTATTAAGTGTATATTTTAATTAACTTCTTAATATATGACC 1620
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      203 AAAATTGTTGATGTCAGGTA 224
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1621 AAAATTGTTGATGTCAGGTA 1642
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 11
US-09-080-625-4          *
; Sequence 4, Application US/09080625
; Patent No. 6307123
; GENERAL INFORMATION:
; APPLICANT: Kriz, Alan L.
; APPLICANT: Spencer, T. Michael
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE
; TITLE OF INVENTION: IDENTIFICATION
; NUMBER OF SEQUENCES: 19

```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/080,625
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, Robert E.
REGISTRATION NUMBER: P-42,628
REFERENCE/DOCKET NUMBER: DEKM161
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3877 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: Linear
US-09-080-625-4

Query Match
Best Local Similarity 97.5%; Score 194; DB 4; Length 3877;
Matches 197; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 23 GGGTGATACGTAAGTCTTCTCTACCTTGATATATATATATATATATATATAT 82
DB 1982 GAGCTGAAGGTAAGTCTCTCTCTACCTTGATATATATATATATATATATAT 2041
QY 83 AGTAGTATATATATATATATATATATATATATATATATATATATATATATAT 142
DB 2042 AGTAGTATATATATATATATATATATATATATATATATATATATATATAT 2101
QY 143 TGCCTTCTGAGTTTATATAGTGTATATATATATATATATATATATATATAT 202
DB 2102 TGCCTTCTGAGTTTATATAGTGTATATATATATATATATATATATATATAT 2161
QY 203 AAAATTGTTGATGTCAGGTA 224
DB 2162 AAAATTGTTGATGTCAGGTA 2183

RESULT 12
US-09-695-782-4
Sequence 4, Application US/09695782
Patent No. 6433252
GENERAL INFORMATION:
APPLICANT: KRITZ, ALAN L.
APPLICANT: GRIFFOR, MATTHEW
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE IDENTIFICATION
FILE REFERENCE: DEKM161USC1
CURRENT APPLICATION NUMBER: US/09/695,782
CURRENT FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: 09/080,625
PRIOR FILING DATE: 1998-05-18
NUMBER OF SEQ ID NOS: 19
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 3877
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
```

```

OTHER INFORMATION: Primer
US-09-695-782-4

Query Match
Best Local Similarity 97.5%; Score 194; DB 4; Length 3877;
Matches 197; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 23 GGGTGATACGTAAGTCTTCTCTACCTTGATATATATATATATATATATATAT 82
DB 1982 GAGCTGAAGGTAAGTCTTCTCTACCTTGATATATATATATATATATATATAT 2041
QY 83 AGTAGTATATATATATATATATATATATATATATATATATATATATATATAT 142
DB 2042 AGTAGTATATATATATATATATATATATATATATATATATATATATATAT 2101
QY 143 TGCCTTCTGAGTTTATATAGTGTATATATATATATATATATATATATATAT 202
DB 2102 TGCCTTCTGAGTTTATATAGTGTATATATATATATATATATATATATATAT 2161
QY 203 AAAATTGTTGATGTCAGGTA 224
DB 2162 AAAATTGTTGATGTCAGGTA 2183

RESULT 13
US-08-817-188-5/C
Sequence 5, Application US/08817188
Patent No. 6074876
GENERAL INFORMATION:
APPLICANT: DE BLOCK, MARC
TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR
FILE REFERENCE: 2121-0127P
CURRENT APPLICATION NUMBER: US/08/817,188
CURRENT FILING DATE: 1997-05-15
EARLIER APPLICATION NUMBER: PCT/EP96/03366
EARLIER FILING DATE: 1996-07-31
EARLIER APPLICATION NUMBER: EP 95401844.6
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 5560
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
OTHER INFORMATION: plasmid pTHW142
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(25)
OTHER INFORMATION: RB: right border sequence of octopine TL-DNA from
OTHER INFORMATION: pT186S3
FEATURE:
NAME/KEY: misc_feature
LOCATION: (84)..(296)
OTHER INFORMATION: 3' 97: 3' untranslated region containing the
OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium
OTHER INFORMATION: T-DNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: (318)..(869)
OTHER INFORMATION: bar: region coding for phosphinothricin
OTHER INFORMATION: acetyltransferase
FEATURE:
NAME/KEY: misc_feature
LOCATION: (830)..(2760)
OTHER INFORMATION: PSSU: promoter region of Rubisco small subunit
OTHER INFORMATION: gene of Arabidopsis thaliana
FEATURE:
NAME/KEY: misc_feature
LOCATION: (2765)..(3058)
OTHER INFORMATION: 3' untranslated region of the CamV 35S transcript
OTHER INFORMATION: containing polyadenylation signals
```

```

? NUMBER OF SEQ ID NOS: 23
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 2
? LENGTH: 1034
? TYPE: DNA
? ORGANISM: Escherichia coli
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (0)...(0)
? OTHER INFORMATION: Glucosamine-6-phosphate deaminase.
US-09-367-293-2

Query Match
Best Local
Matches 190; Similarity 9.4%; Score 190; DB 4; Length 1034;
Mismatch 0; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 33 GTAAGTTTCCTGCTCACTTGATATATAAATAATTATCATTAATTAGTACTATA
DB 235 GTAAAGTTTCCTCTCHACCTTCGTATATATATAATAATTATCATTAATTAGTACTATA
QY 93 TAAATTTTCAAAATTTTTTTCAAAADNAAGAATAGTAAATNCGAATTCCTTTCTGG
DB 295 TAAATTTTCAAAATTTTTTTCAAAADNAAGAATAGTAAATNCGAATTCCTTTCTGG
QY 153 TAATTTAATAGGCGTATATTTTAATTTATATACCTTTCTAATATATAC
DB 355 TAGTTTAAGTGTGATATATTTTAATTTATATACCTTTCTAATATATAC
QY 354

```

213 GATGTCAGG 222  
 DB 415 GATGTCAGG 424  
 RESULT 15  
 US-08-487-826B-13  
 Sequence 13, Application US/08487826B  
 Patent No. 5993827  
 GENERAL INFORMATION:  
 APPLICANT: Shim, Kim L.  
 APPLICANT: Kim, Kim L.  
 APPLICANT: Mills, Chetan  
 APPLICANT: Miller, Louis H.  
 APPLICANT: Patterson, David S.  
 APPLICANT: Su Kin-shaun  
 TITLE OF INVENTION: Wellbms, Thomas E.  
 NUMBER OF SEQUENCES: 45  
 BINDING DOMAINS FROM PLASMODIUM VIVAX  
 CORRESPONDENCE ADDS: 45  
 AND PLASMODIUM FALCIPARUM

STRESSSEE KROBKE ADDRESS:  
STREET: 620 Newport Martens Olson & Bear  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION NUMBER: US/08/487, 826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israetlsen, Ned  
REGISTRATION NUMBER:  
REFERENCE/DOCKET NUMBER: 29,655  
ELECTRONIC COMMUNICATIONS







GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 18:05:33 ; Search time 279.5 Seconds

(without alignments)  
9030.174 Million cell updates/sec

Title: US-09-940-550A-10

Sequence: 1 atggcacaagggtgtgtgac.....cggaagacgtacgagcgtacg 2031

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 783854 seqs, 621352466 residues

Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Published\_Applications\_MA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2031	100.0	2031	9	US-09-940-550A-10
2	2031	99.2	2031	9	US-09-940-550A-9
3	1998	98.4	9801	9	US-09-940-550A-11
4	1795	88.4	1842	10	US-09-788-287-20
5	1790.2	88.1	1842	10	US-09-788-287-26
6	1785.6	87.9	1833	10	US-09-788-297-27
7	1767	87.0	1833	10	US-09-788-297-28
8	195	9.6	2000	10	US-09-887-576-582
9	195	9.6	1180	10	US-09-887-576-581
10	53	2.6	248	10	US-09-969-373-1213
11	51.4	2.5	740	9	US-09-791-279-27
12	50	2.5	14708	9	US-10-239-676-222
13	50	2.5	15577	9	US-10-158-160A-8
14	50	2.5	32367	9	US-10-158-160A-14
15	49	2.4	11036	9	US-10-239-676-117
16	47.8	2.4	6046	9	US-10-239-676-15
17	47.2	2.3	321	10	US-09-960-352-1267
18	47	2.3	5689	9	US-10-239-676-90
19	47	2.3	8996	9	US-10-239-676-212

20	46.6	2.3	11485	9	US-10-104-580-9	Sequence 9, Appl
21	46.4	2.3	428	10	US-09-969-373-1172	Sequence 1172, Ap
22	46.4	2.3	11260	9	US-10-239-676-20	Sequence 20, Appl
23	46.2	2.3	309	10	US-09-919-580-857	Sequence 857, App
24	46	2.3	5979	9	US-10-239-676-18	Sequence 18, Appl
25	46	2.3	6298	9	US-10-239-676-63	Sequence 63, Appl
26	45.6	2.2	89	10	US-09-911-088-3	Sequence 3, Appl
27	45.6	2.2	5908	9	US-10-239-676-94	Sequence 94, Appl
28	45.2	2.2	6298	9	US-10-239-676-64	Sequence 64, Appl
29	45	2.2	426	10	US-09-960-352-6529	Sequence 6529, Ap
30	45	2.2	64081	10	US-09-790-988-1	Sequence 1, Appl
31	44.8	2.2	64081	10	US-09-790-988-1	Sequence 1, Appl
32	44.6	2.2	2559	10	US-09-737-178-109	Sequence 109, Appl
33	44.6	2.2	3002	10	US-09-737-178-119	Sequence 119, App
34	44.6	2.2	11047	9	US-10-239-676-188	Sequence 188, App
35	44.2	2.2	3410	9	US-10-196-063-1	Sequence 1, Appl
36	43.6	2.1	7089	9	US-10-239-676-67	Sequence 67, Appl
37	43.6	2.1	15732	9	US-10-239-676-96	Sequence 96, Appl
38	43.4	2.1	5908	9	US-10-239-676-93	Sequence 93, Appl
39	43	2.1	431	10	US-09-969-373-739	Sequence 739, App
40	43	2.1	472	10	US-09-960-352-368	Sequence 368, App
41	43	2.1	473	10	US-09-969-373-1212	Sequence 1212, Ap
42	43	2.1	11812	9	US-10-239-676-209	Sequence 209, App
43	43	2.1	12405	9	US-10-239-676-36	Sequence 36, Appl
44	43	2.1	127197	9	US-09-754-853A-1	Sequence 1, Appl
45	42.8	2.1	484	9	US-09-918-995-10122	Sequence 10122, A

## ALIGNMENTS

RESULT 1

US-09-940-550A-10

Sequence 10, Application US/09940550A

Publication No. US2003003635A1

GENERAL INFORMATION:

APPLICANT: MANKIN, LUKE

APPLICANT: MCKERSIE, BRYAN

TITLE OF INVENTION: SELF-EXISTING POLYNUCLEOTIDES AND USES THEREOF

FILE REFERENCE: 16313-0055

CURRENT APPLICATION NUMBER: US/09/940, 550A

CURRENT FILING DATE: 2001-08-27

PRIOR APPLICATION NUMBER: 60/227, 961

PRIOR FILING DATE: 2000-08-25

NUMBER OF SEQ ID NOS: 12

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 10

LENGTH: 2031

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: C3Int\* DNA

OTHER INFORMATION: sequence

US-09-940-550A-10

Query Match	100.0%	Score 2031:	DB 9:	Length 2031:
Best Local Similarity	100.0%	Pred. No. 0:		
Matches 2031:	Conservative	0:	Mismatches	0:
			Indels	0:
			Gaps	0:

instaz  
LS

```
Db 181 ATAACTTTCTATATATATGACCAAAATTTGTGATGTGACAGTACGGGGTGTCTACGAC 240
Qy 241 CGTCAGTCGGCGGAGCGCGAATATAGCAGTGCAGCAAGCCCGACGACACAGCTACCGCC 300
Db 241 CGTCAGTCGGCGGAGCGCGAATATAGCAGTGCAGCAAGCCCGACGACACAGCTACCGCC 300
Qy 301 AACGAGACAGAGCGCGGCGGACCTTACGCGGAGTGCAGCGGAGCGGCGGCGGCTTACAG 360
Db 301 AACGAGACAGAGCGCGGCGGACCTTACGCGGAGTGCAGCGGAGCGGCGGCGGCTTACAG 360
Qy 361 TTGCTGGGGGATTTACGCGGAGCGGCGGCGGAGTGCAGCGGAGTGCAGCGGAGTGCAG 420
Db 361 TTGCTGGGGGATTTACGCGGAGCGGCGGAGTGCAGCGGAGTGCAGCGGAGTGCAG 420
Qy 421 GAGTTGGAACGCACTCTGAAACGAAATCCGCGCGGCGCTCAACATGATCTATTTGCTAT 480
Db 421 GAGTTGGAACGCACTCTGAAACGAAATCCGCGCGGCGCTCAACATGATCTATTTGCTAT 480
Qy 481 GAGCTGTCCGCGCTTCTCGCGGCTGAAAGTCTATGACGCGGATTCGATTTGCTCGGAATTG 540
Db 481 GAGCTGTCCGCGCTTCTCGCGGCTGAAAGTCTATGACGCGGATTCGATTTGCTCGGAATTG 540
Qy 541 CTCGCGCTGGGCGTGACGATTTGTTCTCACTACGAAAGGCTTCTCCGCGGAGGAAAGCTC 600
Db 541 CTCGCGCTGGGCGTGACGATTTGTTCTCACTACGAAAGGCTTCTCCGCGGAGGAAAGCTC 600
Qy 601 ATGAGACTTATTCACCTGATTAATGCGGCTGACGCGCTGCACAAAGAAATCTTCGCTGAG 660
Db 601 ATGAGACTTATTCACCTGATTAATGCGGCTGACGCGCTGCACAAAGAAATCTTCGCTGAG 660
Qy 661 TCGGCGAAGATTCCTGACACGAAAGCTTCAGCGCGGAAATTTGGGGGTACGTGCGCGGG 720
Db 661 TCGGCGAAGATTCCTGACACGAAAGCTTCAGCGCGGAAATTTGGGGGTACGTGCGCGGG 720
Qy 721 AAGGCGCTTACGCTTGGAGCTTTTCGAGACGAAAGGAGTACCGCAACGCGCGA 780
Db 721 AAGGCGCTTACGCTTGGAGCTTTTCGAGACGAAAGGAGTACCGCAACGCGCGA 780
Qy 781 ATGTCATATGTCATCAACAAGTTAGCGACTGACACTCCGCTTACCGGACCTTTC 840
Db 781 ATGTCATATGTCATCAACAAGTTAGCGACTGACACTCCGCTTACCGGACCTTTC 840
Qy 841 GAGTTGAGCCCGAGCTAATCCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 900
Db 841 GAGTTGAGCCCGAGCTAATCCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 900
Qy 901 CCTTCAAGCGCGGAGTCAAGCCGCTTATACCGCGGAGCACTACGCGGCTTTTGAAG 960
Db 901 CCTTCAAGCGCGGAGTCAAGCCGCTTATACCGCGGAGCACTACGCGGCTTTTGAAG 960
Qy 961 CGCATGACGCTGACGCGGCTCCGACCGCGGCGGAGAGATTTGGGAAGAACCGCTTCA 1020
Db 961 CGCATGACGCTGACGCGGCTCCGACCGCGGCGGAGAGATTTGGGAAGAACCGCTTCA 1020
Qy 1021 AGCGGCTTGGGAGCGCGGCAACCGTTATGCGAATCTTGGGACCGCGCTTATTTGGGCTTC 1080
Db 1021 AGCGGCTTGGGAGCGCGGCAACCGTTATGCGAATCTTGGGACCGCGCTTATTTGGGCTTC 1080
Qy 1081 GCGGCTGAGGATCTTCAAGAAGAACCGGAGCGGACCGCGGACCGAAGATTTAGAGGT 1140
Db 1081 GCGGCTGAGGATCTTCAAGAAGAACCGGAGCGGACCGCGGACCGAAGATTTAGAGGT 1140
Qy 1141 TACCGCATTCAGCGGACCGGATCACGCTCGCGCGGCTGAGCTTGTATTTGCGGACGATC 1200
Db 1141 TACCGCATTCAGCGGACCGGATCACGCTCGCGCGGCTGAGCTTGTATTTGCGGACGATC 1200
Qy 1201 ATGAGAGCCGCTGAGTGTATGAGCTTACGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1260
Db 1201 ATGAGAGCCGCTGAGTGTATGAGCTTACGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1260
Qy 1261 GGGGCTTCCGCGGAGCAAGCAATTCGTGCGGCATGTGAGACAAAGCTTATGCGAGTGTGCG 1320
Db 1261 GGGGCTTCCGCGGAGCAAGCAATTCGTGCGGCATGTGAGACAAAGCTTATGCGAGTGTGCG 1320

Qy 1321 GCCGTCATGACTTTCGAAGCGCGGAGAGAAATGATCAAGGACTTCTTACCGGTGCCGTCCG 1380
Db 1321 GCCGTCATGACTTTCGAAGCGCGGAGAGAAATGATCAAGGACTTCTTACCGGTGCCGTCCG 1380
Qy 1381 CGGAAGTGTGTGACCCGCTTCCGACCTTGGGACGACGAGAGGACAGTGCAGCTGAGATG 1440
Db 1381 CGGAAGTGTGTGACCCGCTTCCGACCTTGGGACGACGAGAGGACAGTGCAGCTGAGATG 1440
Qy 1441 GCGGCACTGCAAGATTTGTTGCGGAAACGATCTTCAACAGATTCAGACAGCCCGGAAGGC 1500
Db 1441 GCGGCACTGCAAGATTTGTTGCGGAAACGATCTTCAACAGATTCAGACAGCCCGGAAGGC 1500
Qy 1501 GACGAGAGAGCTTGTGCGCTTCTGTGGGAGCCCGGACGCTTGGCAAGCTCAGTACG 1560
Db 1501 GACGAGAGAGCTTGTGCGCTTCTGTGGGAGCCCGGACGCTTGGCAAGCTCAGTACG 1560
Qy 1561 GCGGCTGGAAGAGCGGCGGAGACGCGGCGGAGAGCTTACGAGAGCCGCTTGAAC 1620
Db 1561 GCGGCTGGAAGAGCGGCGGAGACGCGGCGGAGAGCTTACGAGAGCCGCTTGAAC 1620
Qy 1621 GCGCTTGAAGAGCTGTACGAGAGCGGCGGAGAGCTTACGAGAGCCGCTTGAAC 1680
Db 1621 GCGCTTGAAGAGCTGTACGAGAGCGGCGGAGAGCTTACGAGAGCCGCTTGAAC 1680
Qy 1681 AAGCACTTCCGGAAGCAACAGGACGCGCTGACGCTCCGCGAGCAAGAGGCGGGAAGAGCG 1740
Db 1681 AAGCACTTCCGGAAGCAACAGGACGCGCTGACGCTCCGCGAGCAAGAGGCGGGAAGAGCG 1740
Qy 1741 CTTCGCAAGCTTGAAGCGCGGCGGAGAGCCCGGAGTGTGCGGCTTGAACCAATGTGCTCCG 1800
Db 1741 CTTCGCAAGCTTGAAGCGCGGCGGAGAGCCCGGAGTGTGCGGCTTGAACCAATGTGCTCCG 1800
Qy 1801 GAGCGGACGCTGACCGCGGAGCGGCGGAGAGCTTGTGCGGCGGCGGCGGCTGAGTACGAG 1860
Db 1801 GAGCGGACGCTGACCGCGGAGCGGCGGAGAGCTTGTGCGGCGGCGGCGGCTGAGTACGAG 1860
Qy 1861 AAGGCGCTGTTGCTGCGGCGCTTCTGTAAGACAAATGTTGTCAGAAAGTCAAGTACGGGG 1920
Db 1861 AAGGCGCTGTTGCTGCGGCGCTTCTGTAAGACAAATGTTGTCAGAAAGTCAAGTACGGGG 1920
Qy 1921 AAGGCGGAGGAGAAACCGCATGAGAAAGCGGCTTGTATGACGAGTGTGCTGAGTACGAG 1980
Db 1921 AAGGCGGAGGAGAAACCGCATGAGAAAGCGGCTTGTATGACGAGTGTGCTGAGTACGAG 1980
Qy 1981 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2031
Db 1981 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2031

RESULT 2
US-09-940-550A-9
: Sequence 9, Application us/09940550A
: Publication No. US20030033635A1
: GENERAL INFORMATION:
: APPLICANT: MANKIN, LUKE
: APPLICANT: MCKERSIE, BRYAN
: TITLE OF INVENTION: SELF-EXCISING POLYNUCLEOTIDES AND USES THEREOF
: FILE REFERENCE: 16313-0055
: CURRENT APPLICATION NUMBER: US/09/940, 550A
: PRIOR APPLICATION NUMBER: 60/227, 961
: PRIOR FILING DATE: 2000-08-25
: NUMBER OF SEQ ID NOS: 12
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 9
: LENGTH: 2031
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: c3lint DNA
US-09-940-550A-9
```

Query Match 99.2%; Score 2015; DB 9; Length 2031;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 2021; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
QY 1 ATGCACAGGGGTTGTGACCGGGGTGATACGTAAGTTTCTGCTTACCTTTGATATA 60
DB 1 ATGCACAGGGGTTGTGACCGGGGTGATACGTAAGTTTCTGCTTACCTTTGATATA 60
QY 61 TATATATATATATATATATATATATATATATATATATATATATATATATATATAT 120
DB 61 TATATATATATATATATATATATATATATATATATATATATATATATATATATAT 120
QY 121 AAGAATAGTATATATATATATATATATATATATATATATATATATATATATAT 180
DB 121 AAGAATAGTATATATATATATATATATATATATATATATATATATATATATAT 180
QY 181 ATAACTTTTCTAATATATATATATATATATATATATATATATATATATATATAT 240
DB 181 ATAACTTTTCTAATATATATATATATATATATATATATATATATATATATATAT 240
QY 241 CGTAGTCGCGGAGCGGAGAAATAGCAGTACAGAACCCAGCAGCAGTACGCC 300
DB 241 CGTAGTCGCGGAGCGGAGAAATAGCAGTACAGAACCCAGCAGCAGTACGCC 300
QY 301 AACGAAGACAAGGGGCGGACCTTTCAGCGGAGTACAGCGGAGCGGCGGTTCAAG 360
DB 301 AACGAAGACAAGGGGCGGACCTTTCAGCGGAGTACAGCGGAGCGGCGGTTCAAG 360
QY 361 TTGCTGCGGCAATTTTCAGGAAAGCGCGGACGTCGCGGTTTCGGAAGCGCGCCG 420
DB 361 TTGCTGCGGCAATTTTCAGGAAAGCGCGGACGTCGCGGTTTCGGAAGCGCGCCG 420
QY 421 GAGTTCCAGACCATCTCTGAAGAAATGCCGCGCGCGGCGCTACATATATCTAT 480
DB 421 GAGTTCCAGACCATCTCTGAAGAAATGCCGCGCGCGGCGCTACATATATCTAT 480
QY 481 GACGTGTCGCGCTTCTCGCGCTGAAAGTCAATGACGCGATTCGATTTCTCGGAATTG 540
DB 481 GACGTGTCGCGCTTCTCGCGCTGAAAGTCAATGACGCGATTCGATTTCTCGGAATTG 540
QY 541 CTCGCCCTGGGCGTGAAGCAATTTTTCACAGGAAGCGTCTTCGCGAGGAAACGTC 600
DB 541 CTCGCCCTGGGCGTGAAGCAATTTTTCACAGGAAGCGTCTTCGCGAGGAAACGTC 600
QY 601 ATGACCGCATTCACCTGATATATGCGGCTGACGCGTGCACAAAGATTTCTCGTGAAG 660
DB 601 ATGACCGCATTCACCTGATATATGCGGCTGACGCGTGCACAAAGATTTCTCGTGAAG 660
QY 661 TCGGCGAAGATTTCTGACACGAAAGACCTTTCAGCGCGAATTTGGCGGGTACGTCGGCGG 720
DB 661 TCGGCGAAGATTTCTGACACGAAAGACCTTTCAGCGCGAATTTGGCGGGTACGTCGGCGG 720
QY 721 AAGCGCCCTTACGCGCTTTCGAGCTTTGTTTCGAGAGCAAGAGATTCACGCGCAAGCGCGCA 780
DB 721 AAGCGCCCTTACGCGCTTTCGAGCTTTGTTTCGAGAGCAAGAGATTCACGCGCAAGCGCGCA 780
QY 781 ATGTTCAATGTCATCAACAAGTTAGCGACTTCGACACTCCCTTACCGGACCTTTC 840
DB 781 ATGTTCAATGTCATCAACAAGTTAGCGACTTCGACACTCCCTTACCGGACCTTTC 840
QY 841 GAGTTCCAGCGCGAGTAAATCCGCTGCTGCTGCTGATCAAGACGACCAACACCTT 900
DB 841 GAGTTCCAGCGCGAGTAAATCCGCTGCTGCTGCTGATCAAGACGACCAACACCTT 900
QY 901 CCTTTCAAGCGCGGCACTCAAGCGCCATTCACCGGCGACATTCACGCGGCTTTGTAAG 960
DB 901 CCTTTCAAGCGCGGCACTCAAGCGCCATTCACCGGCGACATTCACGCGGCTTTGTAAG 960
QY 961 CGCATGAGCGTGAAGCGCGTCCGACCCGCGGCGAGAGATTTGGGAAAGAACCGCTTCA 1020
DB 961 CGCATGAGCGTGAAGCGCGTCCGACCCGCGGCGAGAGATTTGGGAAAGAACCGCTTCA 1020
```

```
QY 1021 ACGCGCTGGAGCCGCGCAACCTTATGCGAATCTTTCGGAGACCGCGTATTCGCGCTTC 1080
DB 1021 ACGCGCTGGAGCCGCGCAACCTTATGCGAATCTTTCGGAGACCGCGTATTCGCGCTTC 1080
QY 1081 GCCGCTGAGGTATCTTACAGAAGAAAGCGGACGCGACCGCAGCAAGATTTGAGGCT 1140
DB 1081 GCCGCTGAGGTATCTTACAGAAGAAAGCGGACGCGACCGCAGCAAGATTTGAGGCT 1140
QY 1141 TACGCGATTCACGCGGACCGGATACGCGCGCGCGGCTGACACTTATGTCGCGACGATC 1200
DB 1141 TACGCGATTCACGCGGACCGGATACGCGCGCGCGGCTGACACTTATGTCGCGACGATC 1200
QY 1201 ATCGAGCCCGCTGAGTGTATGAGCTTCAGCGGTGTTGAGCGGAGCGGCGCGCAAG 1260
DB 1201 ATCGAGCCCGCTGAGTGTATGAGCTTCAGCGGTGTTGAGCGGAGCGGCGCGCAAG 1260
QY 1261 GGGCTTTCCCGGGGCAAGCCATCTGTCGCCATGAGCAAGCTGTACTGCGAGTGTGCG 1320
DB 1261 GGGCTTTCCCGGGGCAAGCCATCTGTCGCCATGAGCAAGCTGTACTGCGAGTGTGCG 1320
QY 1321 GCCGTCATGACTTTCGAAGCGCGGGAAGATGATCAAGAGCTTTCACGCTGCGCTGCG 1380
DB 1321 GCCGTCATGACTTTCGAAGCGCGGGAAGATGATCAAGAGCTTTCACGCTGCGCTGCG 1380
QY 1381 CGAAGGTGTGACGACCGCTGCGACCTGGGCGAGCAAGAGCGACGTGCAAGCTCAGCATG 1440
DB 1381 CGAAGGTGTGACGACCGCTGCGACCTGGGCGAGCAAGAGCGACGTGCAAGCTCAGCATG 1440
QY 1441 GCGGCACTGACAGATTTGTTGCGGAACGATCTTTCACAGATCAGGCAAGCGCAAGGCG 1500
DB 1441 GCGGCACTGACAGATTTGTTGCGGAACGATCTTTCACAGATCAGGCAAGCGCAAGGCG 1500
QY 1501 GACGAAGAGAGCTTGGCGCTTCTGTGGGAAGCGCGCGCGGCTTGGCGCAAGCTCAGTGA 1560
DB 1501 GACGAAGAGAGCTTGGCGCTTCTGTGGGAAGCGCGCGCGGCTTGGCGCAAGCTCAGTGA 1560
QY 1561 GCGCCTGAGAAAGCGGCGCAACGCGGCAACCTTGTTCGCGAGCGCGCGCAAGCGCTGAAAC 1620
DB 1561 GCGCCTGAGAAAGCGGCGCAACGCGGCGCAACCTTGTTCGCGAGCGCGCGCAAGCGCTGAAAC 1620
QY 1621 GCCCTTGAAGAGCTTACAGAGACCGCGCGGAGAGCTTACAGAGAGCCGCTTGGCAAG 1680
DB 1621 GCCCTTGAAGAGCTTACAGAGACCGCGCGGAGAGCTTACAGAGAGCCGCTTGGCAAG 1680
QY 1681 AAGCACTTCCGGAAGCAAGCAGCGGCTGACGCTCCGCGCAGCAAGGCGGCGGAAGAGCGG 1740
DB 1681 AAGCACTTCCGGAAGCAAGCAGCGGCTGACGCTCCGCGCAGCAAGGCGGCGGAAGAGCGG 1740
QY 1741 CTTCGCCAATTTGAAGCGCGCGCAAGCCCGGAAGTTGCGCTTGAACCAATGTTCCCGCA 1800
DB 1741 CTTCGCCAATTTGAAGCGCGCGCAAGCCCGGAAGTTGCGCTTGAACCAATGTTCCCGCA 1800
QY 1801 GACGCGAGCTGACCGCGACCGCGCTTAACTGTGTGTGGGGGCGCGCTCACTAGACGAC 1860
DB 1801 GACGCGAGCTGACCGCGACCGCGCTTAACTGTGTGTGGGGGCGCGCTCACTAGACGAC 1860
QY 1861 AAGCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1920
DB 1861 AAGCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1920
QY 1921 AAGGGGCGAGGAAAGCCCATCGAGAAGCGCGCTTGTGATCAGCTGGGCGCAAGCGCGCGAC 1980
DB 1921 AAGGGGCGAGGAAAGCCCATCGAGAAGCGCGCTTGTGATCAGCTGGGCGCAAGCGCGCGAC 1980
QY 1981 GACGACGAGAAAGAGAGCGAGCGCCAGAGAGCGGCAAGAGAGCTAGCGCGTAG 2031
DB 1981 GACGACGAGAAAGAGAGCGAGCGCCAGAGAGCGGCAAGAGAGCTAGCGCGTAG 2031
```

RESULT 3  
US-09-940-550A-11  
; Sequence 11, Application us/0994050A  
; Publication No. US20030033635A1

Query Match	98.48;	Score 1998;	DB 9;	Length 9901;
Best Local Similarity	99.28;	Pred. No. 0;		
Matches 2021; Conservative	0;	Mismatches	10;	Indels . 7; Gaps 1

QY	474	TGTCATATACGCGTCGCGCTTCTGCGCGCTGGAAGTCATGAGCGCATTCGCCATTTGCTC	533
Db	5464	TGTCATATACGCTGTCGCGCTTCTGCGCGCTGGAAGTCATGAGCGCATTCGCCATTTGCTC	5522
QY	534	GGATTTGCTCGCCCTGGGGCGTGACGATTTGTTTCCACTCAGGAAGGGCTCTCCGGCAGGG	5933
Db	5524	GGAAATTTGCTCCCGCCGGGCGTGACGATTTGTTTCCACTCAGGAAGGGCTCTCCGGCAGGG	5958
QY	594	AAACGTCATGGAACCGGATTCACCGGATTTATGCGGCTGAGCGCTCGCACAAACAAATCTTC	653
Db	5584	AAACGTCATGGAACCGGATTCACCGGATTTATGCGGCTGAGCGCTCGCACAAACAAATCTTC	564
QY	654	GCTGAAGTCGGCGGAAGATTTTCGACACGAAGAACCCTTCACCGCGAAATTGGCGCGGTACGT	713
Db	5644	GCTGAAGTCGGCGGAAGATTTTCGACACGAAGAACCCTTCACCGCGAAATTGGCGCGGTACGT	570
QY	714	CGGCGGGAAGGGCGGCTTACGGGCTTCGAGCTTTGTTTCGGAGACGAGAGATCAGCGCA	773
Db	5704	CGGCGGGAAGGGCGGCTTACGGGCTTCGAGCTTTGTTTCGGAGACGAGAGATCAGCGCA	576
QY	774	CGGCGGAATGTCATATGTCGTCATCAACAAGTAGGCACTGCAACACATCCCTTCACGG	833
Db	5764	CGGCGGAATGTCATATGTCGTCATCAACAAGTAGGCACTGCAACACATCCCTTCACGG	582
QY	834	ACCCTTGAGATTCGAGCCCGACGTAATCCGGTGCTGGTGCGTGAGATCAAGACSCCAA	893
Db	5824	ACCCTTGAGATTCGAGCCCGACGTAATCCGGTGCTGGTGCGTGAGATCAAGACSCCAA	588
QY	894	ACACCTTCCCTTCAAGCCGGGACATCAAGCCGGCATTTCAACCCGGGACATCAACGGCGCT	953
Db	5884	ACACCTTCCCTTCAAGCCGGGACATCAAGCCGGCATTTCAACCCGGGACATCAACGGCGCT	5943
QY	954	TTGTGAAGCGCATGAGACGCTGACCGCGTGCCGACCCGGGGGAGACGATTTGGAGAAGAAC	1013
Db	5944	TTGTGAAGCGCATGAGACGCTGACCGCGTGCCGACCCGGGGGAGACGATTTGGAGAAGAAC	6002
QY	1014	CGCTTCAAGGCGCCGAGACCCGGCAACCGTTATGGAATTCCTTCGGGACCCCGCATTTGC	1077
Db	6004	CGCTTCAAGGCGCCGAGACCCGGCAACCGTTATGGAATTCCTTCGGGACCCCGCATTTGC	606
QY	1074	GGGCTTGC CGCTGAGGTTGATCTACAAGAAGAAAGCCGAGCGACGCGCACACACAGAT	1133
Db	6064	GGGCTTGC CGCTGAGGTTGATCTACAAGAAGAAAGCCGAGCGACGCGCACACACAGAT	6122
QY	1134	TGAGGGTTACCGGATTCAGGCGCACCCCGATCAACGCTCCGCGGCTGAGCTGATTGGCG	1199
Db	6124	TGAGGGTTACCGGATTCAGGCGCACCCCGATCAACGCTCCGCGGCTGAGCTGATTGGCG	6183
QY	1194	ACCGATCATGACCCCGCTGAGTAGATGATAGATTCAAGGCTGTTTGAACGGCAGGGGGCG	1253
Db	6184	ACCGATCATGACCCCGCTGAGTAGATGATAGATTCAAGGCTGTTTGAACGGCAGGGGGCG	6243
QY	1254	CGGCAAGGGGCTTCCCGGGGGCAGGCCATTCTGTCGCCATGAGCAAGCTGTACTGCGA	1313
Db	6244	CGGCAAGGGGCTTCCCGGGGGCAGGCCATTCTGTCGCCATGAGCAAGCTGTACTGCGA	6303
QY	1314	GTTGTGGCGCGCTATCACTTCGAAAGGCGGGGAAAGATGATCAAGAGACTTTACCGTG	1373
Db	6304	GTTGTGGCGCGCTATCACTTCGAAAGGCGGGGAAAGATGATCAAGAGACTTTACCGTG	6363
QY	1374	CGGTCGCGGAGAGTGCTGACCCCGTCGCACTTGGGGACGACGGAAGGCGCTGCACAGT	1433
Db	6364	CGGTCGCGGAGAGTGCTGACCCCGTCGCACTTGGGGACGACGGAAGGCGCTGCACAGT	6423
QY	1434	CACCATGAGGGGCACTGAGCAAGTGTTGGGGAAGCAATCTTCAACAAGATCAGGCAAGC	1493
Db	6424	CACCATGAGGGGCACTGAGCAAGTGTTGGGGAAGCAATCTTCAACAAGATCAGGCAAGC	6483
QY	1494	CGAAGCGACGAGAGACGTTGGCGCTTCTGAGGAAACCGCGCAGACCTTTCGGCAAGCT	1553
Db	6484	CGAAGCGACGAGAGACGTTGGCGCTTCTGAGGAAACCGCGCAGACCTTTCGGCAAGCT	6543



Db 1344 CCCCCAGCCTTGGCAGACTGAGAGGCGCTTGAAGAGAGCGGCGCAACCT 1403  
Qy 1593 TGTTCGGAGCCGCCGAGCGCCCTTGAAGAGCTGTAGAGAACCGCGCGGC 1652  
Db 1404 TGTTCGGAGCCGCCGAGCGCCCTTGAAGAGCTGTAGAGAACCGCGCGGC 1463  
Qy 1653 AGGACCTTACGACGAGCCGTTGGCAGAGAGCACTTCCGGAAGCAAGCGCTGAC 1712  
Db 1464 AGGCGCTGACGAGCGCTTGGCAGAGAGCACTTCCGGAAGCAAGCGCTGAC 1523  
Qy 1713 GTCGCGGAGCAAGGCGGAGAGAGCGGCTTCCGAGCTTGAAGCGCGCGCA 1772  
Db 1524 GTCGCGGAGCAAGGCGGAGAGAGCGGCTTCCGAGCTTGAAGCGCGCGCA 1583  
Qy 1773 GTTCCGCTTGAACATGTTCCCGAAGAGCGCGCTGACCGCGAGCGCTTAAGTC 1832  
Db 1584 GCTTCCCTTGAACATGTTCCCGAAGAGCGCGCTGACCGCGAGCGCTTAAGTC 1643  
Qy 1833 GTGTTGGGGGGCGCGGTAGTAGAGAGAGCGCGCTTGTCTCGGGCTTCTAGACAA 1892  
Db 1644 GTGTTGGGGGGCGCGGTAGTAGAGAGAGCGCGCTTGTCTCGGGCTTCTAGACAA 1703  
Qy 1893 GATCCTTGTACAGAGTACTAGCGGAGGAGGAGGAGCGCGCATCGAAGCGCGC 1952  
Db 1704 GATCCTTGTACAGAGTACTAGCGGAGGAGGAGGAGCGCGCATCGAAGCGCGC 1763  
Qy 1953 TTCGATCAGTGGGAGAGCGCGGAGCGAGCGAGAGAGAGAGCGCGAGCGCAC 2012  
Db 1764 TTCGATCAGTGGGAGAGCGCGGAGCGAGCGAGAGAGAGAGCGCGAGCGCAC 1823  
Qy 2013 GGAAGAGTAGCGGCGGTAG 2031  
Db 1824 GGAAGAGTAGCGGCGGTAG 1842

RESULT 5  
US-09-788-297-26  
Sequence 26, Application US/09788297  
Patent No. US20020094516A1  
GENERAL INFORMATION:  
APPLICANT: Calos, Michele P.  
APPLICANT: Scilment, Christopher R.  
TITLE OF INVENTION: ALTERED RECOMBINASES FOR GENOME MODIFICATION  
FILE REFERENCE: 8400-0011  
CURRENT APPLICATION NUMBER: US/09/788,297  
CURRENT FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 26  
LENGTH: 1842  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: altered  
OTHER INFORMATION: recombinase 7C1  
US-09-788-297-26

Query Match 88.1%; Score 1790.2; DB 10; Length 1842;  
Best Local Similarity 99.0%; Pred. No. 0;  
Matches 1801; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 213 GATGTGAGGTACGCGGGTCTTACGACGCTAGTCGCGGAGCGGAGAAATAGCAGTCG 272  
Db 24 GTGTGACACGTACGCGGGTCTTACGACGCTAGTCGCGGAGCGGAGAAATTCGAGCGC 83  
Qy 273 AGGAAGCCGAGCAACAGCTAGTACGCCCAAGCAAGAGCGCGCGCTTACGCGCA 332  
Db 84 AGGAAGCCGAGCAACAGCTAGTACGCCCAAGCAAGAGCGCGCGCTTACGCGCA 143  
Qy 333 AGTCAGCGGAGCGGCGGCTTACGCTTGTCTCGGGCAATTCAGGAGAGCGCGCGCAC 392  
Db 144 AGTCAGCGGAGCGGCGGCTTACGCTTGTCTCGGGCAATTCAGGAGAGCGCGCGCAC 203

Qy 393 GTCCGCGCTTGGGAGCGGCGGAGCGCGGAGTTCGAAGCATCTCTGAACCAATGCCCGC 452  
Db 204 GTCCGCGCTTGGGAGCGGCGGAGCGCGGAGTTCGAAGCATCTCTGAACCAATGCCCGC 263  
Qy 453 CGGCGCGCTCAACATGATTCATTTGTATGACGTGTCCGCTTCCGCGCTGAGAGTCAT 512  
Db 264 CGGCGCGCTCAACATGATTCATTTGTATGACGTGTCCGCTTCCGCGCTGAGAGTCAT 323  
Qy 513 GGAGCGCATTCGCAATTTGTCTCGGATTTCTCGGCTTGGCGGTGACGATTTCTTCCATCA 572  
Db 324 GGAGCGCATTCGCAATTTGTCTCGGATTTCTCGGCTTGGCGGTGACGATTTCTTCCATCA 383  
Qy 573 GGAAGGCGCTTCCGCGGAGGAAACGTCATGAGCTGATTCACCTGATTTAGCGGCTCGA 632  
Db 384 GGAAGGCGCTTCCGCGGAGGAAACGTCATGAGCTGATTCACCTGATTTAGCGGCTCGA 443  
Qy 633 CGCGTCCACAAAGATCTTCCGTGAAGTGGGCGAAGATTCGACAGCAAGAACCTTCA 692  
Db 444 CGCGTCCACAAAGATCTTCCGTGAAGTGGGCGAAGATTCGACAGCAAGAACCTTCA 503  
Qy 693 GCGGATTTGGGCGGCTACGTCGCGGAGGAGCGCGCTTACGCGCTTGTTCGCA 752  
Db 504 GCGGATTTGGGCGGCTACGTCGCGGAGGAGGAGCGCGCTTACGCGCTTGTTCGCA 563  
Qy 753 GACGAAGAGATCAGCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 812  
Db 564 GACGAAGAGATCAGCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 623  
Qy 813 CTGAGCAGCTCCCTTACGCGGAGCGCTTCAAGTTCGAGCGCGAGCGATCCGCTGCTG 872  
Db 624 CTGAGCAGCTCCCTTACGCGGAGCGCTTCAAGTTCGAGCGCGAGCGATCCGCTGCTG 683  
Qy 873 GCGTGAATCAAGAGCAGCAACACCTTCCCTTCAAGCGCGGAGGAGGAGGAGGAGGAGGAG 932  
Db 684 GCGTGAATCAAGAGCAGCAACACCTTCCCTTCAAGCGCGGAGGAGGAGGAGGAGGAGGAG 743  
Qy 933 CCGGCGAGCATCAGGCGGCTTGTAAAGCGATGAGCGCGCTGAGCGCGAGCGCGGAG 992  
Db 744 CCGGCGAGCATCAGGCGGCTTGTAAAGCGATGAGCGCGCTGAGCGCGAGCGCGGAG 803  
Qy 993 CGAGAGATTTGGGAGAGAACCGCTTCAAGCGCTTGGAGCGCGGAGCGCGGAGCGCGGAG 1052  
Db 804 CGAGAGATTTGGGAGAGAACCGCTTCAAGCGCTTGGAGCGCGGAGCGCGGAGCGCGGAG 863  
Qy 1053 CCTTCGGAGCGCGCTTGTAAAGCGATGAGCGCGCTGAGCGCGAGCGATCAAGAGAGCGGAG 1112  
Db 864 CCTTCGGAGCGCGCTTGTAAAGCGATGAGCGCGCTGAGCGCGAGCGATCAAGAGAGCGGAG 923  
Qy 1113 CGGAGCGCGGAGCGAGAGATGAGGCTTACGCGATTCAGCGCGAGCGCGATCAGCGCTCG 1172  
Db 924 CGGAGCGCGGAGCGAGAGATGAGGCTTACGCGATTCAGCGCGAGCGCGATCAGCGCTCG 983  
Qy 1173 GCGGCTGAGCTTGAATTTGGGAGCGATCAGCGCGCTGAGCGCGAGCGCGATCAGCGCTCG 1232  
Db 984 GCGGCTGAGCTTGAATTTGGGAGCGATCAGCGCGCTGAGCGCGAGCGCGATCAGCGCTCG 1043  
Qy 1233 GTGTTGGAGCGGAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1292  
Db 1044 GTGTTGGAGCGGAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1103  
Qy 1293 CATGGCAAGCTGTACTGAGTGTGGCGGCTGATGACTTGAAGCGCGGAGGAGGAGGAGGAG 1352  
Db 1104 CATGGCAAGCTGTACTGAGTGTGGCGGCTGATGACTTGAAGCGCGGAGGAGGAGGAGGAG 1163  
Qy 1353 GATCAAGAGCTTTCACGCTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1412  
Db 1164 GATCAAGAGCTTTCACGCTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1223  
Qy 1413 GCAGCAAGGAGCTGCAAGCTGACGATGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1472  
Db 1224 GCAGCAAGGAGCTGCAAGCTGACGATGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1283  
Qy 1473 CTTCAAGAGATCAGGAGCGGCGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1532





Qy	1413	GCACGAGGCGACGTGTCAACGTCACCAATGTGGGGCGCACTGACAGTGTGTTGCCGAACGCAT	1472
Db	1224	GCACGAGGCGACGTGTCAACGTCACCAATGTGGGGCGCACTGACAGTGTGTTGCCGAACGCAT	1283
Qy	1473	CTTCAACAAGATCAGAGCGACGCGCGAAGCGAGCAAGAGACGTTGGCGCTTCTGTGGGAAGC	1532
Db	1284	CTTCAACAAGATCAGAGCGACGCGCGAAGCGAGCAAGAGACGTTGGCGCTTCTGTGGGAAGC	1343
Qy	1533	CGCCGCGACGCTTGGCGAAGCTCACTGAGCGCCTGAGAAAGCGGGCGAAGCGGCGACACT	1592
Db	1344	CGCCGCGACGCTTGGCGAAGCTCACTGAGCGCCTGAGAAAGCGGGCGAAGCGGCGACACT	1403
Qy	1593	TGTTGCGGAGGCGCGCGACGCGCTGGAAGCGCCTTGAAGAGCTGTACGAAGACCGCGCGGC	1652
Db	1404	TGTTGCGGAGGCGCGCGACGCGCTGGAAGAGCTGTACGAAGACCGCGCGGC	1463
Qy	1653	AGGACCTTACGACGAGACCGCTTGGCAGGAAGCACTTCGGAAGCAACAGCGAGCGTAC	1712
Db	1464	AGGACCTTACGACGAGACCGCTTGGCAGGAAGCACTTCGGAAGCAACAGCGAGCGTAC	1523
Qy	1713	GCTCGGAGGAGGAAGGGCGGGAAGAGCGGCTTGCCGAACCTTGAAGCGCGCGAAGCCCGAA	1772
Db	1524	GCTCGGAGGAGGAAGGGCGGGAAGAGCGGCTTGCCGAACCTTGAAGCGCGCGAAGCCCGAA	1583
Qy	1773	GTTGCCCTTGACCAATGTTTCCCGGAAGACGCGCAGCTGACCGCAGCGCCCTTAAGTC	1832
Db	1584	GTTGCCCTTGACCAATGTTTCCCGGAAGACGCGCAGCTGACCGCAGCGCCCTTAAGTC	1643
Qy	1833	GTTGGGGGGGGCGCGCGCTATTAACAGCAACAGCGCGTGTTCGTGGGGCTTTCGTAGACAA	1892
Db	1644	GTTGGGGGGGGCGCGCGCTATTAACAGCAACAGCGCGTGTTCGTGGGGCTTTCGTAGACAA	1703
Qy	1893	GATCGTTGTACAGAAAGTCAGACTAGGGGCGAGGGGCGAGGGAGCGCCATCAGGAAGCGCGC	1952
Db	1704	GATCGTTGTACAGAAAGTCAGACTAGGGGCGAGGGGCGAGGGAGCGCCATCAGGAAGCGCGC	1763
Qy	1953	TTCGATCAGCTGGGCGAAGCGCGCGACGAGCAGCAGCAAGCAAGCAAGCGCCCGAGCGCAC	2012
Db	1764	TTCGATCAGCTGGGCGAAGCGCGCGACGAGCAGCAGCAAGCAAGCAAGCGCCCGAGCGCAC	1823
Qy	2013	GGAAGACGTAGCGGC 2028	
Db	1824	GGAAGACGTAGCGGC 1839	
RESULT 7			
US-09-788-297-28			
: Sequence 28, Application us/09788297			
: Patent No. US20020094516A1			
: GENERAL INFORMATION:			
: APPLICANT: Calos, Michele P.			
: APPLICANT: Solimanti, Christopher R.			
: TITLE OF INVENTION: ALTERED RECOMBINASES FOR GENOME MODIFICATION			
: FILE REFERENCE: 8400-0011			
: CURRENT APPLICATION NUMBER: US/09/788,297			
: CURRENT FILING DATE: 2001-02-16			
: NUMBER OF SEQ ID NOS: 34			
: SOFTWARE: PatentIn Ver. 2.1			
: SEQ ID NO 28			
: LENGTH: 1833			
: TYPE: DNA			
: ORGANISM: Artificial Sequence			
: FEATURE:			
: OTHER INFORMATION: Description of Artificial Sequence: altered			
US-09-788-297-28			

[illegible]

D	24	GGTGGACAGCTAGCGGGGTGCTTAGACACCGTACGTGCGCGGACGCCGAATAATTGAGCGC	83
Q	273	AGCAAGCCCAAGCAGCACACCGTATAGCGCCACGAGAAGAGGGGGCGCACTTTCAGCGCGA	333
D	84	AGCAACCCCAAGCAGCACACCGTATAGCGCCACGAGAAGAGGGGGCGCACTTTCAGCGCGA	144
Q	333	AGTCGAGCCGAGACGGGGGGCGGGTTCAGAGTTCGTGGGGATTTTCAGCGGAAGCCCGGGCAC	392
D	144	AGTCGAGCCGAGACGGGGGGCGGGTTCAGAGTTCGTGGGGATTTTCAGCGGAAGCCCGGGCAC	203
Q	393	GTTCGCGTTTCGGAGACGGCGGAGCGGCCGAGTTTCGAACGCATCTCTGAACGAATGCTCCGC	452
D	204	GTTCGCGTTTCGGAGAGGGCGGAACGCCCGGAGTTTCGAACGCATCTCTGAACGAATGCTCCGC	262
Q	453	CGGGGGCGCTCAACATGATCATCTGTCATGACGTGTGCGGCTTCTCGCGCTTGAAGTTCAT	512
D	264	CGGGGGCGCTCAACATGATCATCTGTCATGACGTGTGCGGCTTCTCGCGCTTGAAGTTCAT	323
Q	513	GGACGCGATTCGGAATGTCTCGGAATTTCTCGCCCTGGGGCTGACGATTTGTTTCACCTCA	572
D	324	GGACGCGATTCGGAATGTCTCGGAATTTCTCGCCCTGGGGCTGACGATTTGTTTCACCTCA	383
Q	573	GGAAAGCGCTTTCGGGGACGGGAAGCGTATGACCTGATTCACCTGATTTATGCGGCTCGA	632
D	384	GGAAAGCGCTTTCGGGGACGGGAAGCGTATGACCTGATTCACCTGATTTATGCGGCTCGA	443
Q	633	CGCGTCGACAAAGAAATCTTGCTGGAAGTGGCGGACAGTTTCTGACACGGAAGAAACCTTCA	692
D	444	CGCGTCGACAAAGAAATCTTGCTGGAAGTGGCGGACAGTTTCTGACACGGAAGAAACCTTCA	503
Q	693	GGCGCAATTTGGGGGGGTATGCTCGGGCGGGAAAGCGCCTTACGGCTTCGAGCTTGTTCGGA	752
D	504	GGCGCAATTTGGGGGGGTATGCTCGGGCGGGAAAGCGCCTTACGGCTTCGAGCTTGTTCGGA	563
Q	753	GACGAAGAGATACAGCGCCACGGCCGAATGTGCTAAATGTGTCATACAAAGTTACGCA	812
D	564	GACGAAGAGATACAGCGCCACGGCCGAATGTGCTAAATGTGTCATACAAAGTTACGCA	623
Q	813	CTCGACCACTCCCTTACCGGACCCCTTCGAGTTTCGAGCCCGACGATATCGGTGGTGTG	872
D	624	CTCGACCACTCCCTTACCGGACCCCTTCGAGTTTCGAGCCCGACGATATCGGTGGTGTG	683
Q	873	GGGTAGATCAAGACGGCAAAACCTTCCCTTCAAGCCGGGACATCAAGCCGCCATTTCA	932
D	684	GGGTAGATCAAGACGGCAAAACCTTCCCTTCAAGCCGGGACATCAAGCCGCCATTTCA	743
Q	933	CCCGGGGACATCAAGGGGGCTTGTAAAGCCATGAGAGCGTGAAGCGCTCCGACCCGGGG	992
D	744	CCCGGGGACATCAAGGGGGCTTGTAAAGCCATGAGAGCGTGAAGCGCTCCGACCCGGGG	803
Q	993	CGAAGCATTTGGGAAGAAGACCGCTTCAAGCGCTTGAAGCCGGGACCCGGCAACCGTTATGCAAT	1052
D	804	CGAAGCATTTGGGAAGAAGACCGCTTCAAGCGCTTGAAGCCGGGACCCGGCAACCGTTATGCAAT	863
Q	1053	CGTTTGGGAGACCGGGGTTCGCGGGCTTCGCGGCTTGAAGGATCTCAAAAGAAAGCGGA	1112
D	864	CGTTTGGGAGACCGGGGTTCGCGGGCTTCGCGGCTTGAAGGATCTCAAAAGAAAGCGGA	923
Q	1113	CGGACGCCCAACAGCAAGATTAGGGTTACCGCATTTACGCGAGACCCGATCACGCTCG	1172
D	924	CGGACGCCCAACAGCAAGATTAGGGTTACCGCATTTACGCGAGACCCGATCACGCTCG	983
Q	1173	GGCGGTGACCTTGATTGGGAGCCGATTCATACAGCCCGGTGAGTGTATGACCTTCAGC	1233
D	984	GGCGGTGACCTTGATTGGGAGCCGATTCATACAGCCCGGTGAGTGTATGACCTTCAGC	1043
Q	1233	GTGGTTGGAGCGAGGGGGCGCGCAAGGGGGCTTTCGCGGGGCAAGCATTTCTGCGC	1292
D	1044	GTGGTTGGAGCGAGGGGGCGCGCAAGGGGGCTTTCGCGGGGCAAGCATTTCTGCGC	1103
Q	1293	CATGAGCAAGCTTACTCGGATGTGTGGCGCCTCATGACTTTCGAAGCGCGGGGAAGAATC	1352
D	1104	CATGAGCAAGCTTACTCGGATGTGTGGCGCCTCATGACTTTCGAAGCGCGGGGAAGAATC	1163



```
QY 1353 GATCAAGACTCTTACCGCTGCCGCGAAGGTGTGACACCGTCCGACCTTGCGCA 1412
      |||
Db 1164 GATCAAGAGCTTTACCGTGGCGCGGGAAGGTGTGACACCGTCCGACCTTGCGCA 1223
      |||
QY 1413 GCACGAAGGACGTCAACGTGACATGGCGGCTCTGACAAAGTTCGTCGGAACGCAT 1472
      |||
Db 1224 GCACGAAGGACGTCAACGTGACATGGCGGCTCTGACAAAGTTCGTCGGAACGCAT 1283
      |||
QY 1473 CTTCACAAAGATCAGGACGCGGGAAGGCGGCAAGAGAGAGAGAGAGAGAGAGAG 1532
      |||
Db 1284 CTTCACAAAGATCAGGACGCGGGAAGGCGGCAAGAGAGAGAGAGAGAGAGAGAG 1343
      |||
QY 1533 CGCCCGAGCTTCGCGCAAGCTCAGTACAGGCGCTGAGAGAGAGAGAGAGAGAGAG 1592
      |||
Db 1344 CGCCCGAGCTTCGCGCAAGCTCAGTACAGGCGCTGAGAGAGAGAGAGAGAGAGAG 1403
      |||
QY 1593 TGTTCGCGAGCGCGCGGAGCGGCTTGAAGAGCTGTACGAAGAGAGAGAGAGAGAG 1652
      |||
Db 1404 TGTTCGCGAGCGCGCGGAGCGGCTTGAAGAGCTGTACGAAGAGAGAGAGAGAGAG 1463
      |||
QY 1653 AGGACCTTAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1712
      |||
Db 1464 AGGACCTTAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1523
      |||
QY 1713 GCTCCGCGAGCAAGGAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1772
      |||
Db 1524 GCTCCGCGAGCAAGGAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1583
      |||
QY 1773 GTTGGCCCTTGACCAATGTTCCCGAAGACGCGGACGCTGACCGGCGGCTTAAGTC 1832
      |||
Db 1584 GCTTCCCTTGACCAATGTTCCCGAAGACGCGGACGCTGACCGGCGGCTTAAGTC 1643
      |||
QY 1833 GTGGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1892
      |||
Db 1644 GTGGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1703
      |||
QY 1893 GATCGTTGTACAGAGTCACTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1952
      |||
Db 1704 GATTTGTGTACAGAGTCACTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1763
      |||
QY 1953 TTGCATCAGTGGGCGAAGCCCGGACGACGACGACGACGACGACGACGACGACGAC 2012
      |||
Db 1764 TTGCATCAGTGGGCGAAGCCCGGACGACGACGACGACGACGACGACGACGACGAC 1822
      |||
QY 2013 GGAAGACGTAG 2023
      |||
Db 1823 GGAAGACGTAG 1833
      |||
```

RESULT 8  
US-09-887-576-582

```
Sequence 582, Application US/09887576
Patent No. US20020144047A1
GENERAL INFORMATION:
APPLICANT: Budworth, P.
APPLICANT: Brown, D.
APPLICANT: Chang, H.
APPLICANT: Han, B.
APPLICANT: Wang, X.
APPLICANT: Cooper, Bret
TITLE OF INVENTION: Promoters for regulation of plant expression
FILE REFERENCE: 1360.001US1
CURRENT FILING DATE: 2001-06-25
PRIOR APPLICATION NUMBER: US 60/213,848
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: US 60/214,087
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: US 60/258,692
PRIOR FILING DATE: 2000-12-29
NUMBER OF SEQ ID NOS: 875
```

```
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 582
LENGTH: 2000
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-887-576-582

Query Match
Best Local Similarity 100.0%; Pred. No. 1,7e-36;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 TACGTAAGTTTCGCTTACCTTGAATATATATATATATATATATATATATATATAT 89
      |||
Db 383 TACGTAAGTTTCGCTTACCTTGAATATATATATATATATATATATATATATATAT 442
      |||
QY 90 ATATATATTTTCAAAATTTTTCACAAATTAAGAAATAGATATATAGCAATTCCTTT 149
      |||
Db 443 ATATATATTTTCAAAATTTTTCACAAATTAAGAAATAGATATATAGCAATTCCTTT 502
      |||
QY 150 CTGTAGTTTATAGTGTATATTTTATATATATATATATATATATATATATATATAT 209
      |||
Db 503 CTGTAGTTTATAGTGTATATTTTATATATATATATATATATATATATATATATAT 562
      |||
QY 210 GTTGATGTGACGTA 224
      |||
Db 563 GTTGATGTGACGTA 577
      |||
```

RESULT 9  
US-09-887-576-581

```
Sequence 581, Application US/09887576
Patent No. US20020144047A1
GENERAL INFORMATION:
APPLICANT: Budworth, P.
APPLICANT: Brown, D.
APPLICANT: Chang, H.
APPLICANT: Han, B.
APPLICANT: Wang, X.
APPLICANT: Cooper, Bret
TITLE OF INVENTION: Promoters for regulation of plant expression
FILE REFERENCE: 1360.001US1
CURRENT APPLICATION NUMBER: US/09/887,576
CURRENT FILING DATE: 2001-06-25
PRIOR APPLICATION NUMBER: US 60/213,848
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: US 60/214,087
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: US 60/258,692
PRIOR FILING DATE: 2000-12-29
NUMBER OF SEQ ID NOS: 875
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 581
LENGTH: 11180
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-887-576-581
```

```
Query Match
Best Local Similarity 100.0%; Pred. No. 4.7e-36;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 TACGTAAGTTTCGCTTACCTTGAATATATATATATATATATATATATATATATAT 89
      |||
Db 2231 TACGTAAGTTTCGCTTACCTTGAATATATATATATATATATATATATATATATAT 2290
      |||
QY 90 ATATATATTTTCAAAATTTTTCACAAATTAAGAAATAGATATATAGCAATTCCTTT 149
      |||
Db 2291 ATATATATTTTCAAAATTTTTCACAAATTAAGAAATAGATATATAGCAATTCCTTT 2350
      |||
QY 150 CTGTAGTTTATAGTGTATATTTTATATATATATATATATATATATATATATATAT 209
      |||
Db 2351 CTGTAGTTTATAGTGTATATTTTATATATATATATATATATATATATATATATAT 2410
      |||
```



QY	103	AAATATTTTCAAAATTAAGAAATGCTAGTATATATGCAATGCTTCTGAGTTTAA	162
DB	11248	GCAGCGTTTTCGTAATATATATATTTTAACTTTTTCGATATATTTATATATTA	11307
QY	163	GTGCTATATTTTAAATTTATATACCTTTCTAATAT	196
DB	11308	TTTATTTTAAATTTTATATATATTTTATATTTT	11341
RESULT 13			
US-10-158-160A-8/c			
: Sequence 8, Application US/10158160A			
: Publication No. US2003005805A1			
: GENERAL INFORMATION:			
: APPLICANT: RAPPOLD-HOERBRAND, GUDDRUN			
: APPLICANT: RAO, ESCOLE			
: TITLE OF INVENTION: HUMAN GROWTH GENE AND SHORT STATURE GENE REGION			
: FILE REFERENCE: 108351-00004			
: CURRENT APPLICATION NUMBER: US/10/158,160A			
: CURRENT FILING DATE: 2002-08-20			
: PRIOR APPLICATION NUMBER: 09/147,699			
: PRIOR FILING DATE: 1999-06-24			
: PRIOR APPLICATION NUMBER: PCP/EP97/05355			
: PRIOR FILING DATE: 1997-09-29			
: PRIOR APPLICATION NUMBER: 60/027,633			
: PRIOR FILING DATE: 1996-10-01			
: PRIOR APPLICATION NUMBER: EP/97100583.0			
: PRIOR FILING DATE: 1997-01-16			
: NUMBER OF SEQ ID NOS: 55			
: SOFTWARE: PatentIn Ver. 2.1			
: SEQ ID NO 8			
: LENGTH: 15577			
: TYPE: DNA			
: ORGANISM: Homo sapiens			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (3844)..(4068)			
: OTHER INFORMATION: PET92 region (first part)			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (4326)..(4437)			
: OTHER INFORMATION: PET92 region (second part)			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (4545)..(4619)			
: OTHER INFORMATION: PET92 region (third region)			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (5305)..(5512)			
: OTHER INFORMATION: part of exon II (ET93)			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (1498)..(1807)			
: OTHER INFORMATION: part of exon I (G310)			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (11620)..(11729)			
: OTHER INFORMATION: part of exon IV (G108)			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (2665)			
: OTHER INFORMATION: a, c, t, g, other or unknown			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (2880)			
: OTHER INFORMATION: a, c, t, g, other or unknown			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (3508)			
: OTHER INFORMATION: a, c, t, g, other or unknown			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (3529)			
: OTHER INFORMATION: a, c, t, g, other or unknown			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (3555)			
: OTHER INFORMATION: a, c, t, g, other or unknown			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (3578)			
: OTHER INFORMATION: a, c, t, g, other or unknown			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (3580)			
: OTHER INFORMATION: a, c, t, g, other or unknown			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (3592)			
: OTHER INFORMATION: a, c, t, g, other or unknown			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (3600)			
: OTHER INFORMATION: a, c, t, g, other or unknown			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (3606)			
: OTHER INFORMATION: a, c, t, g, other or unknown			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (3609)			
: OTHER INFORMATION: a, c, t, g, other or unknown			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (3619)			
: OTHER INFORMATION: a, c, t, g, other or unknown			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (3621)			
: OTHER INFORMATION: a, c, t, g, other or unknown			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (3643)			
: OTHER INFORMATION: a, c, t, g, other or unknown			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (3681)..(3682)			
: OTHER INFORMATION: a, c, t, g, other or unknown			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (3729)			
: OTHER INFORMATION: a, c, t, g, other or unknown			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (3867)			
: OTHER INFORMATION: a, c, t, g, other or unknown			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (3965)			
: OTHER INFORMATION: a, c, t, g, other or unknown			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (4174)			
: OTHER INFORMATION: a, c, t, g, other or unknown			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (4212)			
: OTHER INFORMATION: a, c, t, g, other or unknown			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (4376)			
: OTHER INFORMATION: a, c, t, g, other or unknown			



```

?      2000-04-07
?
?      2000-06-30
?      2000-09-01
?      NUMBER OF SEQ ID NOS: 228
?      SEQ ID NO 117
?      LENGTH: 11036
?      TYPE: DNA
?      ORGANISM: Artificial Sequence
?      FEATURE:
?      OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
?      FEATURE:
?      NAME/KEY: unsure
?      LOCATION: (1937..1928, 1943, 1956, 1958, 1981)
?      US-10-239-676-117

```

Query Match	2.4%	Score 49	DB 9	Length 11036
Best Local Similarity	56.5%	Pred. NO	0.3%	
Matches 91	Conservative 0	Mismatches 70	Indels 0	Gaps 0

[illegible]

Search completed: May 11, 2003, 23:31:48  
Job time : 391.5 secs

**This Page Blank (uspto)**